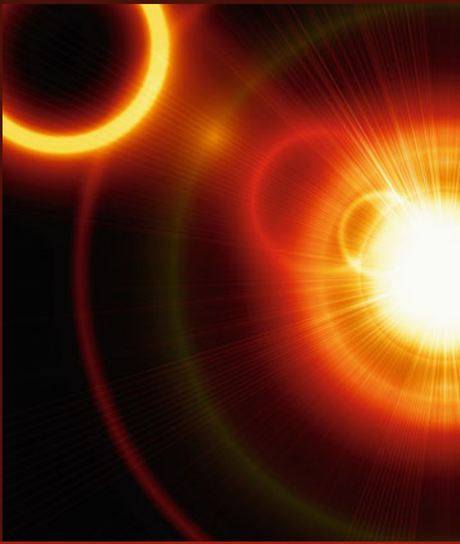
THE FRONTIERS COLLECTION



Richard T. W. Arthur

THE REALITY OF TIME FLOW

Local Becoming in Modern Physics

Springer

THE FRONTIERS COLLECTION

THE FRONTIERS COLLECTION

The books in this collection are devoted to challenging and open problems at the forefront of modern science and scholarship, including related philosophical debates. In contrast to typical research monographs, however, they strive to present their topics in a manner accessible also to scientifically literate non-specialists wishing to gain insight into the deeper implications and fascinating questions involved. Taken as a whole, the series reflects the need for a fundamental and interdisciplinary approach to modern science and research. Furthermore, it is intended to encourage active academics in all fields to ponder over important and perhaps controversial issues beyond their own speciality. Extending from quantum physics and relativity to entropy, consciousness, language and complex systems—the Frontiers Collection will inspire readers to push back the frontiers of their own knowledge.

More information about this series at http://www.springer.com/series/5342

For a full list of published titles, please see back of book or springer.com/series/5342

Richard T. W. Arthur

The Reality of Time Flow

Local Becoming in Modern Physics

Springer

Richard T. W. Arthur
Department of Philosophy
McMaster University
Hamilton, ON, Canada

ISSN 1612-3018 ISSN 2197-6619 (electronic)
THE FRONTIERS COLLECTION
ISBN 978-3-030-15946-7 ISBN 978-3-030-15948-1 (eBook)
https://doi.org/10.1007/978-3-030-15948-1

Library of Congress Control Number: 2019935153

© Springer Nature Switzerland AG 2019
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated with gratitude to my Argentine mentors, Simon Altmann (1924–) in Oxford, and Mario Bunge (1919–) in Montreal, for setting me on my way.

Preface

Time is a topic of perennial interest. We never have enough of it, it waits for no one, and flows on relentlessly. In the calmness of reflection, though, we seem to be able to remove ourselves from its dominion, and to contemplate things as they really are in their essence. The objects of thought and contemplation seem suspended from time.

This ability we have to reflect on things, and especially to view ourselves in abstraction from the relentless fray around us, is part of the essence of consciousness. But the opposition between the temporal (the domain of coming to be and passing away) and the eternal (the world as revealed in contemplation) has also been the foundation of religion and mysticism. In fact, it would seem that as far back as we can trace evidence of self-consciousness in hominids (not just homo sapiens), it has been bound up in religious symbolism. That, however, is supposed to have been before we became wise, emerging into enlightenment from our "self-imposed tutelage", as Kant put it. Since the Enlightenment, we have outgrown the need to found our most general theories about nature on an imagined world of gods and spirits. Now we have science, and through the impartial observation and objective reasoning that science promotes, we can understand the changes and processes occurring in the natural world, at least in principle. We may not know the origins of life, but Darwin taught us how to understand the processes by which life forms have evolved from (ultimately) more primitive organisms. We know how diseases are spread, about the processes by which cells multiply and animals reproduce. In geology, we have learned how even the continents have come to have their present form as a result of the flow of magma and the grinding processes of plate tectonics. And we may not know the origins of the cosmos, but in modern cosmology we understand a great deal about how the galaxies, stars and planets formed, and the laws by which the cosmos developed from its fiery beginnings. In short, the world revealed by science is one of evolution, development, and processes, large and small.

From this point of view, it is perhaps surprising to learn that the majority of modern theoretical physicists and scientifically informed philosophers side with the mystics and idealists, and hold that the passing of time is an illusion, an effect of how we engage a world that in its inner reality is wholly static. Such is the view urged on us by most theoretical physicists, especially those attempting to construct a theory of quantum gravity, uniting the two most successful theories of contemporary physics, General Relativity and Quantum Mechanics. Moreover, they have been encouraged to hold that time flow is illusory by some of the most eminent philosophers of physics, who hold that time in itself has no direction and that talk of time flow is senseless. The physicists in question are well informed about the history of their subject, and open to philosophical debate. And the philosophers in question are knowledgeable about the relevant physics. This makes the need to show where both groups have gone wrong all the more pressing, and that is the point of this book.

The project for the book had its origins some ten years ago in conversations around a table in the Boulangerie Première Moisson on Sherbrooke St. in Montreal in 2008. We were attending a conference there (on The Nature and Ontology of Spacetime), and Steve Savitt (Philosophy, UBC) and I were having a pleasant lunch with Angela Lahee of the Physics Editorial Department of Springer. On hearing (with some surprise) that Steve and I dissented from the widely held view that the passing of time is incompatible with modern physics, Angela immediately suggested that the two of us (and possibly others) should write a monograph on the topic. Dennis Dieks of Utrecht came on board, and the upshot was the acceptance by Springer in late September 2009 of a proposal for a book titled *The Now in Physics*, with each of us (Dennis, Steve and myself) contributing two chapters each. We got a couple of chapters written but then, as is often the way with ambitious joint projects, our various other projects and obligations got in the way of our bringing it to completion. Not wanting to see the project lost, in early 2013 I suggested going it alone, and got a new contract with Springer for this book in early 2015.

Since it is widely believed by physicists and philosophers alike that time's passing is excluded in a fundamental description of physical reality, the book should appeal to a wide audience. It is intended to make a forceful contribution to current philosophical debate about the nature of time. It also has direct bearing on the issue of the incompatibility between relativity theory and quantum theory in their treatments of time. Since this is a central issue for physicists trying to forge a theory of quantum gravity, it constructively engages with those endeavours. Accordingly it is written in such a way as to be readily understandable by physicists and philosophers, students in these disciplines, and a wider readership of interested non-academics. My ideal lay reader would perhaps be someone who reads articles in *Scientific American* or *New Scientist*, but is sometimes frustrated by their lack of philosophical depth; or just someone fascinated by the issues, who wants to learn how they are supposed to be supported by modern physical theory.

I am hopeful that my arguments can be followed by physicists with no training in philosophy, although there is no doubt that the going will get tough in parts. I make no apology for this. Philosophers of physics have taken seriously the need to learn physics to a sufficient depth in order to comment on its philosophical commitments or shortcomings, as they perceive them. But many physicists have all too often not reciprocated and have given free rein to philosophical speculations about topics like determinism, free will, time travel, consciousness and the mind-body problem without first trying to appreciate some of their philosophical intricacies and difficulties.

The book is also intended for an audience of philosophers. Here I have tried to steer a middle course, so that much of it will be comprehensible by those without detailed knowledge of modern physics, but so that it also engages philosophers of physics, even if it is not presented with the same degree of technical sophistication as is normal in their publications.

The strong emphasis I have given to the historical development of thought on time is intended as instructional for both groups: for analytic philosophers who would treat, say, the “B theory”, without troubling to understand either its origin in British idealism or the developments in mathematics or physics that prompted it; but also for physicists who attribute all of classical physics to Newton, without any awareness of the contributions of Huygens or Leibniz, or how they paved the way for Einstein’s views on relativity.

I am indebted to numerous people for their generous feedback on my drafts. Chapter 2 benefited from the critiques of Dennis Dieks and Steve Savitt when it was a contribution to the original co-authored book, and I am grateful to both of them for their continued support and encouragement, as well as for constructive comments on individual chapters. A great deal of what I have to say represents the original common ground we had all found ourselves to agree on, but I have probably gone beyond our core agreement in many places. I am grateful to David Wright for his enthusiastic critical responses throughout the process of composition, to Barry Allen for critical remarks that showed a perfect comprehension of my intentions, and to Steve for his careful reading of the whole manuscript. To Ralph Pudritz, Cliff Burgess and Jon Stone I owe the impetus provided by our co-taught course on the Origins of Spacetime. Lauren Naraine and Sean Dudley gave me welcome feedback from students’ perspectives on the beginning chapters, and I am very grateful to Dennis, Jo Edwards, Mauro Dorato and Carlo Rovelli for the expert input that they found time to give me nearer the end. In between, Roberto Torretti gave me encouragement for some of my heresies in Chaps. 4 and 8, and David Hyder gave me valuable criticisms of my presentation of Einstein in Chap. 6. Other debts are of longer standing and harder to recall. My views on relativity and quantum theory have been nourished by a continued dialogue over the last two decades with Kent Peacock—some basic ideas of Chap. 6 were originally conceived in part as a prompt to a collaborative project with him—as well as by many stimulating conversations with (among others) Jim Brown, Arun Bala, Malcolm Forster, Vesselin Petkov, Mélanie Frappier, Lee Smolin, Bill Unruh, Yuval Dolev, Harvey Brown, Oliver Pooley and Wayne Myrvold. Last but not least has been the constant support of my wife and severest critic, Gabriella.

What I offer here I have mostly learned from others, although none of them is responsible for the way I have presented it. If I have stumbled upon anything original, let us hope the originality is not like that of the errors on a student’s logic exam! As Borges wrote, “A lucky line here and there should not make us think any higher of ourselves, for such lines are the gift of Chance or the Spirit; only the errors are our own.” (Borges 1998, 332).

Hamilton, Canada Richard T. W. Arthur

Reference

Borges, Jorge Luis. 1998. *Collected Fictions*. Transl. Andrew Hurley. New York: Penguin.

Contents

Chapter 1
Introduction

Yet becoming exists; it is a fact.
—Bergson, *Creative Evolution* (1944, 343).

Time is not a physical *thing* like a river. Nor is time itself a *process* that can flow at a certain rate, since if this is taken literally we would need a further time against which to measure this rate. The point is, though, that time flow should *not* be taken literally in either of these senses. When we speak of the flow of time or compare it to a river, we are employing a metaphor. Time does not actually *do* anything, but anything that *does* do something, does so in time. This is what we call the passage of time. It alludes to the fact that processes lead inexorably onwards from the past toward the future, so that, for instance, a motion is not only a passage over a certain space, but also a passage through a certain time. *It is time flow in this sense of passage, the reality of temporal becoming*, that I will be defending in this book.

Now, it might seem that this is an odd thing to be defending. What could be more obvious than the fact that time passes? The most obvious things, however, are often those that are the most difficult to explain. What could be more obvious than that heavy objects tend to fall to the ground? But look at the history of attempts to explain that! Aristotle argued that things have natural places depending on their composition. Those things containing mostly earth tend towards the centre of the cosmos, and settle in places below those mainly consisting of water (these being the two heavy elements), whereas those made of air and fire tend to rise up towards the heavens. This at any rate gave explanations for the position of the Earth at the centre of the cosmos, its being covered with oceans and atmosphere, and even of its spherical shape. Displacing Earth from the centre was no easy matter, requiring a whole new physics, culminating in Newton's theory of gravity as a force acting instantaneously between any two bodies, such as the Earth and his famous apple, inversely proportional to the square of the distance between them and proportional to their masses. But how could one understand such a force? Newton rested his

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_1

case on the fact of the matter, in the face of charges of unintelligibility from his contemporaries, and his physics prevailed because of its astonishing success. Einstein was not satisfied, however, and with his General Theory of Relativity, which we will meet in these pages, he gave an even more counter-intuitive explanation: bodies tend to fall because they are following a "geodesic", that is an optimal path through a spacetime that is itself curved by the presence in it of heavy masses like the Earth. One can go further, and say that the reason things fall is that "they naturally incline toward where time passes more slowly" (Rovelli 2018, 12), since time passes more slowly toward the centre of the Earth, as we'll see in Chap. 7.

The case with the passing of time itself is somewhat different. Unlike the fact of falling bodies, temporal becoming has been regarded with suspicion from the start. No sooner had rational thought established itself in ancient Ionia, replete with explanations of the evolution of the cosmos and of mankind, than it turned against the idea of time's passing. In Parmenides' philosophy change is demoted to mere appearance, and so it is with Plato. As the latter wrote in his *Timaeus* in the 4th century BC,

> What is it that always is and has no coming into being, and what is it that is always coming into being but never is? The one is to be grasped by the mind with reason and is always in the same state. The other is opined by opinion combined with irrational sense perception, and keeps coming into being and going out of existence, but never has real being. (Plato, *Timaeus*, 27D6; quoted from Sorabji 1983, 111)

This emphasis on timeless form over changing matter runs deep in Western thought. And here I am not just referring to the castigation of the body in Christian doctrine,[1] but the tendency in physics to equate reality with mathematics, especially with equations—a trend evident in Hawking's idea that all reality could be represented in an equation on a T-shirt, and perhaps reaching its apotheosis in Max Tegmark's preposterous hypothesis that the physical world is nothing but an abstract mathematical structure.[2]

But we are not just dealing with an ideology. There is a whole battery of arguments that have been trained on the idea of temporal becoming since Plato's time: there is only a manifold of events, it is said, and the temporal relations among these never change; moreover, it is impossible for one point-event to come out of a preceding one since there are no two point-events next to one another in the temporal continuum; the very idea of events changing their relations to the 'now' is fraught with contradiction; temporal becoming is an illusion, a subjective impression, that has its origin in the fact of increasing disorder along one of the two possible directions of time; passage is refuted by the fact that according to special relativity, there is no possibility of

---

[1] This aspect of Plato's philosophy was of course due to his being influenced by Christianity, if we are to believe some of my time-disoriented former students!

[2] In an interview in *New Scientist* (18 November, 2006), Tegmark endorsed Hawking's prediction: "In 50 years, you may be able to buy T-shirts on which are printed equations describing the unified laws of our universes". This cult of equations is also well-illustrated by Neil Turok's "master equation" (Turok 2012, 167–176, illustrated in the centrefold), and his interpretation of Raphael's famous painting as depicting Pythagoras "absorbed in writing equations in a big book" (52)—almost two millennia before Islamic scholars had invented the very idea of equations!

an advancing world-wide now; again, in general relativity objective time lapse is shown to be an illusion by the fact that there is no distinguished time coordinate, and that in a curved spacetime you could (in principle) time travel into your own past; and finally it is claimed that the price of making general relativity compatible with quantum theory is a recognition that time itself is unreal.

Don’t worry if these objections are not yet clear or leave you with questions; I shall be explaining them in detail in what follows. The important thing to recognize at this point is that they are serious objections. There is nothing frivolous about them, and answering them properly will help us to a more profound understanding not only of time and things temporal, but also of the interpretation and implications of some of our deepest and most revolutionary theories of physics. The matters to be discussed are also very subtle. I can make this point by reference to a beautifully written book by the physicist Carlo Rovelli, a book whose title makes it seem like a rejoinder to this one: *Reality is Not What it Seems* (2017). This impression will be strengthened by the knowledge that one of Rovelli’s main contentions is that “Time Does Not Exist” (the title of his Chap. 7), and that “Time plays no role at the fundamental level of physics” (249). And yet in certain respects his views are not far from those I will be advancing here, as evidenced by his favourable references to time’s passage—“time on Earth passes more quickly at higher altitude” (85), “time passes differently in different places” (150). One of the things he is opposing is the idea of an overall “container time”, within which everything can be conceived as happening, with a unique present moment advancing ever onwards. I agree about that: the time that passes, I shall argue here, is a local time, the rate at which processes evolve, and which indeed “passes differently in different places”.

Rovelli claims that in fundamental physics time is eliminated altogether, so he would no doubt object to my taking literally his claims about time’s passing, as would many of his peers among modern physicists. Bertrand Russell had already declared in 1903 that recent advances in the foundations of mathematics showed conclusively that “we live in an unchanging world”, and that change reduces to a mere difference in states at different times (Russell 1903, 347). A similar view has been held to result from Einstein’s overturning of our classical preconceptions about time. Brian Greene summarizes it well: “In fact, a reframing of some of Einstein’s insights from special relativity provides evidence that time does not flow” (Greene 2004, 130). Every time-slice of spacetime, he argues, “exists on the same footing as every other, suggesting, as Einstein believed, that reality embraces past, present and future *equally* and that the flow we envision bringing one section to light as another goes dark is illusory” (132).

Greene and Rovelli are among the most eloquent and talented expositors of modern physics. If it therefore gives me pangs of regret to be taking issue with what they have to say on my subject, this is no less true of Julian Barbour. His *Discovery of Dynamics* (2001) is a beautifully written book, planned as the first historical part of a two-volume work on a Machian approach to motion, space and time. The second volume was never written, but in it, Barbour had expected, “time would play a large role”. That is a very Pickwickian statement, since he reveals in the same passage that on the interpretation of general relativity he was to develop there, “time and duration arise within that theory from an arena in which there is no time at all" (xi). This timeless basis of General Relativity, he continues, has "profound implications for attempts to make it compatible with quantum mechanics"—in other words, for the creation of a theory of quantum cosmology in which "time will cease to play any role in the foundations of physics" (xi–xii). This is in fact a statement of the bold program that Barbour will prosecute over the following years. As we shall see in Chap. 8, Barbour's Machian programme is to build up the universe out of "places", "where 'place' means a relative arrangement, or configuration, of the complete universe" (Barbour 1999, 69). In quantum theory a "configuration" is a representation of the state of the universe at an instant, and Barbour calls the totality of all possible relative configurations "Platonia". But "there is nothing outside the universe to time it as it goes from one place to another in Platonia" (69). Reminiscently of Russell, change is simply difference of one relative configuration from another, and passage is an illusion.

Not all modern physicists and philosophers of physics agree with Rovelli and Barbour, of course. But most of the premises on which their views are based are generally accepted. These have deep roots in thinking about time by both philosophers and physicists, whose negative conclusions have tended to reinforce one another. This explains how I have set about tackling the issues in this book. I start with the deepest roots, finding one in an argument for the unreality of time that is still current, even though two and a half millennia old. I then work my way through other arguments against passage, mainly originating in the nineteenth- and early twentieth centuries, finishing with those arguments with the shallowest roots, arising in certain recent approaches to quantum gravity. In between I do a lot of explaining of the philosophical nuances and arguments from physics that have motivated these attacks on the reality of becoming.

Thus in Chap. 2 I analyze an argument against the reality of time reported by Aristotle, endorsed by many philosophers through the years since. This argument turns on an an equivocation between two ways in which events can be said to exist or be real: they can exist now, that is, at the time I am making this statement, or they can be events that are real in the sense that they actually do occur at the times of their occurrence, as opposed to remaining mere possibilities. The same equivocation, I maintain, undermines the modern doctrine of *presentism*, according to which all that exists, exists now. But there is an analogous ambiguity in the rival doctrine of *eternalism*, which asserts the reality of spacetime and all the events contained in it, to the exclusion of becoming. I argue that events exist neither timelessly nor at all times, but at the times of their occurrence, assuming they occur when they do.

Presentism and eternalism are often identified with the 'A' and 'B' theories of time, respectively, a dichotomy originating with John McTaggart's analysis in his 1908 essay "The Unreality of Time". The 'B' theory, however, is a description of the theory of time that Russell had given earlier in his *Principles of Mathematics* of 1903. There Russell proposes a "static" theory of change, explicitly endorsing the argument against becoming and change given by Zeno of Elea in Ancient Greece. This is part of his "at-at theory of motion", which he claims follows from a correct analysis of the continuum—a source of the animus toward becoming that has been largely ignored by recent philosophers. I argue that if Russell's argument against the reality of passage were successful, it would equally undermine the reality of motion.

In Chap. 3 I pursue these theses into modern philosophy of time. Here I examine what prompted Henri Bergson's *cri de coeur* against the spatialization of time, echoed by Alfred North Whitehead. There is, I believe, a strong tendency among physicists to believe that time is simply a fourth dimension of space, a tendency already evident in Bergson's contemporaries, such as Hinton and H. G. Wells. But as Russell observed, spatialization does not consist in representing the order of time as a linear order, but in mistaking properties of the representation of a thing for properties of the thing represented. That applies, ironically enough, to Russell's theory itself: that events are statically represented as having become in a certain order does not entail that they exist independently of their coming to be in that order. The same fallacy, conversely, can be seen to underlie continuing attempts to inject dynamism into physics by having spacetime as a whole subject to change, in order to head off criticisms of it as an unchanging block or static manifold. But spacetime as a whole is neither unchanging nor changing, neither growing nor remaining the same, since time is included as one of its four dimensions.

I also defend Whitehead's contentions that events, as they exist in fact, are always extended in time, and that the point-events of physics are useful abstractions made by shrinking them to the limit. As a consequence, they inherit the temporal directionality of process. The assertion of such a temporal directionality of process is a key contention of my analysis of classical physics in Chap. 4, and an issue on which I part company with the great majority of physicists and philosophers. I argue that concrete processes—as opposed to types of process, of which they are tokens—are necessarily directed from past to future. They are instances of passage from an initial state to a final state. The direction of time is simply this (local) passage of processes from past to future, a reversible process being a type of process that can be oriented either way round with respect to the direction of passage. No one thought to deny this before the intervention of Boltzmann at the end of the nineteenth century, with his claim that the direction of time is given by the direction of increasing entropy. Nowadays, however, it seems to be routinely accepted that there is no direction of time at the micro-level, and that it emerges only at the macrolevel as an effect of entropy. This ignores the doubts that have been raised from the beginning about the cogency of Boltzmann's attempt to derive the Second Law of Thermodynamics governing thermodynamically irreversible processes from mechanical laws that are time-symmetric. At the root of this difficulty, I suggest in Chap. 4, is the widely held conviction that the source of irreversibility is to be sought in laws alone, and not in the statistical improbability of initial conditions used in conjunction with the laws. In addition, though, one may question the applicability of Boltzmann's analysis to the universe as a whole, the feasibility of composing temporally directed processes from undirected ones, and the very cogency of defining the direction of time in terms of increasing entropy in the first place.

In the fifth chapter I turn to the contention (summarized by Greene above) that the relativity of simultaneity in special relativity refutes the flow of time. I maintain that such claims take for granted that the coordinate time in relativity tracks the rate of passage for processes, whereas in fact the rate of change of processes is tracked by proper time, a new concept introduced by Hermann Minkowski. What is crucial about this development is that it allows the notion of becoming to be separated from a conception of it as occurring by the advance of a global ‘now’, a kind of hyperplane on which future is converted to past. Once this idea of becoming as the passage of a global now has been put aside, I contend, there is nothing to impugn the idea of passage of time as a fundamental *local* feature of temporal reality. Becoming, on this view, is simply the coming about of events from other events in their neighbourhood, over a non-zero period of time. As I shall argue, the idea of process as taking a finite time to go from an initial state to a final state (without any necessity for the former to cause the latter) may be used as a founding assumption (along with the principle of relativity and some reasonable symmetry assumptions) to derive the Lorentz transformations that form the core of the special theory of relativity. In general relativity, the same principle of Lorentz invariance must hold for any process at least locally, this requirement being enshrined in Einstein’s Equivalence Principle.

This still leaves open the problem of how to account for the immediacy of our experience of what is present. It is perhaps a majority opinion that, given the relativity of simultaneity, there is no place for the ‘now’ in physics, that the present is something subjective, having to do with the way reality is experienced but not with the way reality is. I contest this in Chap. 6, and show how to define a serviceable notion of the present as objective and relative to a segment of a world line. As I argue, the doctrine of the subjective now has drawn significantly on a misinterpretation of Einstein’s special relativity as entailing that time is relative to the observer. But it is crucially important to distinguish what an observer in a given state of motion might *observe* to be now, and what the same observer might *infer* to be now, as I show using the famous “Twin Paradox” as an illustration. Time dilation—the relativistic running slow of one clock moving relative another—is relative, but not merely apparent. It will result in objective differences in the proper time of two processes that have taken different paths through spacetime from one point to another. Time lapses at different rates along different paths through spacetime.

These differences in the rate of flow are accentuated in general relativity, where gravity also slows the passage of time. This slowing of time flow for processes closer to the source of a gravitational field is, like the time dilation of the special theory, a relative but real effect. The light emitted by a distant star is at a lower frequency (is “red-shifted”) relative to us, because time flows more slowly where it was emitted, closer to the centre of the star’s gravitational field. But no change of reference frame alters the facts about which bodies are sources of gravitation, and which are moving inertially. This is important because inertial motions embody the standard for rate of flow: the shortest distances in spacetime, the “geodesics”, are those along which the proper time is maximized.

All of this presumes the existence of events and of the processes they constitute. But the assumption of relativity theory that there exist processes with definite trajectories in spacetime raises some serious problems of interpretation in quantum theory. For example, the naïve model of an atom as a nucleus surrounded by orbiting electrons, like planets orbiting the Sun, cannot be sustained: the “electron cloud” around the nucleus is a cloud of probability, the probability of its being found in a certain location, or with a certain energy. Where there are interactions of a certain kind, there are events—for instance, when an electron hits a screen and produces a flash of light. But the equations of quantum theory do not predict such events, only their probabilities of occurring. This has led to some extravagant interpretations: that the observer brings phenomena into being by observing them; that all possibilities whose probabilities the theory predicts are actual in some universe or other; that there is no unique trajectory connecting the present with some particular past, just records or memories of a such past that exist in the present. It is also often claimed that quantum correlations involve non-local influences. In Chap. 8 I sketch the main features of quantum theory and the difficulties in its interpretation that have given rise to such interpretations, and subject them to criticism.

In the final section of the book I turn to the question of time in quantum gravity, where the conclusions reached in previous chapters (as well as in this chapter) all have some relevance. I argue that the arguments of Barbour and Rovelli purporting to show the inevitability of the elimination of time, variously turn on assumptions whose legitimacy I have previously questioned: the assumption that the appearance of becoming could be accounted for without presupposing the becoming of the appearance (Chap. 2); the idea that since all events are already included in the model, they simply *are*, and do not need also to “become” (Chap. 3); the presupposition that the present of consciousness is instantaneous (Chaps. 3 and 6); the notion that time direction is an emergent phenomenon (Chap. 4); the misconception that the relativity of simultaneity (or the covariance of the equations of general relativity) threatens the objectivity of becoming (Chaps. 5 and 7); the idea that the present depends on the observer (Chap. 6); the assumption that equations governing the large scale structure of the cosmos are appropriate for treating becoming in local processes (Chap. 7); and, in common with the many worlds interpretation of quantum theory, the assumption of a wave function of the whole universe, and the conceit that probabilities, and indeed events, can be satisfactorily accounted for using just configurations or relative states (Chap. 8).

That describes the negative side of my argument, the explosion of the many misconceptions and dogmas that purport to prove the unreality of becoming. But there are corresponding positive theses, and I would not want them to go unmentioned. So let me attempt a summary of the overall picture.

There are events everywhere in spacetime, and each of these is a short process, and therefore an instance of becoming. Although strictly speaking there is no becoming in an instant, nor is there becoming from one instant to the next (assuming the denseness property of the continuum, where there are no two instants without a third between them), there are still processes of becoming from one instant to others in its neighbourhood. At any point in spacetime there are events that have occurred in its local past, and events that are about to occur in its immediate future. But there is no “God’s-eye view”, a perspectival point from which all events may be conceived as happening now or having happened already. There is no such thing as the (unique) class of all those events that are happening now, although we may still define a local region of spacetime that contains those events that could be called present to us during a given period of proper time, such as the time it takes us to be conscious of them.

There are local currents of becoming everywhere, because there are processes everywhere. In classical physics the rate of change of all processes is gauged by absolute time. As we shall see, this universal standard of equable flow is embodied in the equable motion of bodies undergoing inertial motions. This gauging of the equable flux of time is taken over into special relativity theory, where inertial motions still have a privileged status. In the context of the general theory of relativity, time flows at different rates in different parts of spacetime, more slowly where the gravitational field is stronger. Spacetime is curved, and so are the trajectories of bodies undergoing inertial motion, the geodesics. Through Einstein's geodesic principle, however, these inertial motions still encapsulate the standard of the equable flow of time by comparison with which all the other local rates of flow are gauged.

Finally, on the question of becoming in quantum theory, quantum probabilities can be taken as representing tendencies to manifest or actualize, tendencies that are indeterministic. It is certainly the case that some of these indeterministic tendencies issue in actual outcomes, in irreversible interactions with systems in their environment, in events. Whatever the links and entanglements among such tendencies prior to their actualization, they always actualize locally. Becoming in quantum theory is thus the local actualization of tendencies.

References

Barbour, Julian B. 1999. *The End of Time: The Next Revolution in Physics*. (2001 reprinting). Oxford: Oxford University Press.

Barbour, Julian B. 2001. *The Discovery of Dynamics*. [An edited reissue of *Absolute or Relative Motion? Vol. 1: The Discovery of Dynamics* (Cambridge: Cambridge University Press, 1989)]. Oxford: Oxford University Press.

Bergson, Henri. 1944. *Creative Evolution*. (English translation of *L'Évolution créatrice*, 1907). New York: Henry Holt and Company (originally published 1911).

Greene, Brian. 2004. *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*. New York: Alfred A. Knopf.

McTaggart, J.M.E. 1908. The unreality of time. *Mind* 17: 457–474.

Rovelli, Carlo. 2017. *Reality Is Not What It Seems: The Journey to Quantum Gravity*. New York: Riverhead Books.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Russell, Bertrand. 1903. *Principles of Mathematics*. New York: W. W. Norton & Co.

Sorabji, Richard. 1983. *Time, Creation, and the Continuum: Theories in Antiquity and the Early Middle Ages*. Ithaca, New York: Cornell University Press.

Tegmark, Max. 2006. Max Tegmark Forecasts the Future. *New Scientist* 192 (2578), November 18.

Turok, Neil. 2012. *The Universe Within: From Quantum to Cosmos*. Toronto: House of Anansi Press.

Chapter 2
The Problem of Time in Classical Philosophy

Without motion we could not perceive the passage of time.
—Isaac Barrow, *Geometrical Lectures* ([1670] 1916, 35).

2.1 Introduction

In this chapter I will be reviewing some of the long-standing issues about time and its passage that are relevant to the treatment of time in modern physics. This will take us into some territory that will be unfamiliar to most physicists, and indeed some of the historical content may be unfamiliar to many modern philosophers of science too. Part of the point of beginning with such a historical chapter is to provide valuable context for the discussion to follow. But in following these rich veins of traditional thought about time we will also find many precedents for contemporary views, including some deep-lying confusions about time and its passing that have persisted into contemporary thought. These matters in Chaps. 3 and 4, where we will examine their manifestations in contemporary philosophy and in the interpretation of classical physics. It is therefore important to try to identify the classical precedents that have set the scene for these contemporary controversies before moving on in later chapters to a treatment of the more recondite issues as they arise in relativity theory and quantum theory.

2.2 Difficulties Concerning the Reality of Time and Passage

Reflection about the nature of time seems to date from the early sixth century BCE, when thinkers in Asia Minor began the process of emancipating cosmology from the anthropomorphism characterizing earlier mythical thought. This can be seen in the reported writings of such thinkers as Pherecydes of Syros (whom Aristotle described

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_2

as writing in a semi-mythical vein), and his contemporary, Anaximander of Miletos (c. 610–c. 546 BC), the first thinker in whose writings we see rational arguments given for cosmological theses and genealogies. In mythical thought, time (*chronos* in Greek) is often portrayed as a god.[1] Pherecydes gives us glimmerings of a justification for this kind of view: time is divine because it is uncreated. Also, it is regarded as a kind of generative principle, since it is through time that everything that is created comes into existence.[2] In one of the very few fragments of Anaximander's thought that have been preserved, time dictates a kind of "arrangement", compensating for the opposing tendencies of things to come to be and pass away at each others' expense.[3] Within a mere two centuries of these beginnings, though, the Greeks had already achieved a high level of sophistication in thought about time. For by the time of Aristotle (384-322 BCE), students in the burgeoning schools were occupying themselves with such abstract problems as the reality of time, the status of the 'now', and the continuity of time.

Thus when Aristotle began his discussion of time with a summary of the main difficulties that needed addressing, the latter were three of the main difficulties about time that he recorded. First he reported considerations that would make one suspect "either that it does not exist at all, or at least that its existence is tenuous and faint" (*Physics* iv, 217 b32).[4] Then he proceeded to questions about the status of "το νυν", the 'now' or present moment: is the now always the same thing, or is it always different? Either supposition seems to lead to paradox. If it is always different, a now which has ceased to exist must have ceased to exist at some earlier now, but two different nows cannot be simultaneous. If it is always the same, then earlier and later events will be occurring at the same now, and "nothing would be either earlier or later than anything else" (*Physics* iv, 218 a8-29). Third, there are puzzles raised by Aristotle's claim that the 'now' is an indivisible boundary separating past from future, and at the same time binding them into a continuous whole. This led many authors—including Diodorus Cronus shortly after Aristotle, Islamic theologians in

---

[1] A fully mythic treatment of time can be found in the *Atharva Veda*, composed before 1000 BCE (although not committed to writing until after the time of Pherecydes and Anaximander): "Time carries us forward, a steed, with seven rays, a thousand eyes, undecaying, full of fecundity. … Time hastens onward, the first god. … It is he who drew forth the worlds, and encompassed them. Being the father, he became their son." (*Atharva Veda XIX*, 53; Muir 1861, 408).

[2] Diogenes Laertius quotes Pherecydes as saying that "Zas [Zeus] and Chronos [Time] always existed, and so did Chthonie [Earth]…". Put simply, nothing could exist before time, since 'before' presupposes time. According to Damascius, Pherecydes held Zas, Chronos and Chthonie to be the three first principles, and asserted that "Time from his own seed created fire and air and water" (Barnes 1987, 58). The idea that time is a kind of generative principle is not absent from contemporary thought. As we shall see in what follows, some physicists insist on regarding time as a generative principle producing new nows, and set about trying to give a physical explanation of such production.

[3] Anaximander is reported to have said that "the things from which existing things come into being are also the things into which they are destroyed, … *for they give justice and reparation to one another for their injustice in accordance with the arrangement of time*". (Barnes 1987, 75). See Carlo Rovelli's *Anaximander* (2007) for an excellent account of this early thinker's profound importance.

[4] All quotations from Aristotle are from (Aristotle 1984), unless otherwise noted.

the eighth century, and several Cartesians in the seventeenth—to claim that time consists of indivisibles or time atoms, so that it is not in fact continuous.[5]

Intriguingly, these roughly correspond to three of the main difficulties concerning time that are of interest to modern physicists and philosophers. Thus Julian Barbour and Carlo Rovelli have claimed that the way forward for reconciling the two great theories of modern physics, quantum theory and relativity, is to acknowledge the *unreality of time* that is signalled by its absence in the fundamental equations of modern physics. Regarding the second difficulty mentioned by Aristotle, the great majority of modern physicists and philosophers have concurred that there is *no room for the* '*now*' in the modern physical worldview. Like 'here', 'now' is not something that features in the equations of physics. The events of spacetime are said to be all equally real, so the classical idea of reality coming into being by one set of events occurring 'now', to be succeeded by another set of events becoming at a later 'now', seems not to feature in such theories.[6] As we shall see, these difficulties are further compounded by the relativity of simultaneity in Einstein's special theory of relativity where there are no "world-wide instants", to use Eddington's memorable turn of phrase (Eddington 1929, 47). Philosophers, meanwhile, have had their own reasons for being sceptical about the idea of a moving now. Movement, they object, surely presupposes time, so that the movement of the 'now' from earlier to later seems to presuppose another dimension of time, and to begin an impossible infinite regress of times.[7]

Concerning the question of the continuity of time, again there is scepticism among both physicists and philosophers. It has been suggested by proponents of each of the two main approaches to a theory of quantum gravity, String Theory and Loop Quantum Gravity, that a theory of gravity consistent with quantum theory will require us to reject the continuity of time.[8] This would require becoming to occur in discrete steps, if it occurs at all. Some philosophers, such as Whitehead, have argued for something similar. For the majority of modern philosophers, however, the difficulty is not with continuity itself, but with attempts to construe becoming in terms of a continuous transition from one instantaneous state to another. The idea that passage requires an instantaneous tendency to change state, represented by Newton's "fluxions" or Leibniz's infinitesimal differences, was revived by the neo-Kantian philosopher Hermann

---

[5] See (Sorabji 1983) for an engaging and scholarly account of theories of time, creation and the continuum in antiquity and the early middle ages. His book is especially valuable for its accounts of Aristotle's views, his treatment of time atomism in antiquity, and his comparisons of these theories with modern views.

[6] "The universe", writes Jack Smart, "is a four-dimensional space-time manifold. Present, past and future are all equally real." (Smart 1968, 255); similarly the physicist Paul Davies: "all events—past, present and future—are equally real" (Davies 1995, 260).

[7] It should be noted that not everyone has regarded multiple times, or even an infinite regress of times, as impossible. We will return to the issue of multiple dimensions of time in Chap. 3 below.

[8] Thus using String Theory one cannot probe geometry below the Planck scale, and this has been taken to indicate that such geometry does not exist at that scale. The Loop Quantum Gravity approach is based on the premise that there is no background spacetime metric, and one of its implications is the discreteness of areas and volumes at the Planck scale. See (Rovelli 1998, 9). But this is too complex an issue to tackle in this book.

Cohen in the nineteenth century. But it was met with scathing rebuttal by Bertrand Russell in the early 1900s, who insisted that infinitesimals had been banished from mathematics by the theory of the continuum developed by Weierstrass, Dedekind and Cantor. Motion, he contended, consists in being in one state at one time, and a later state at a later time, with no need to presume any such thing as a passage from one to the other. Modern philosophers of science have mostly agreed: the understanding of time as involving passage from one state to another is generally rejected in favour of such an austere "static" view.

In this book I claim that when the passage of time is correctly understood, there is nothing in modern physics or mathematics to impugn it. But to make that claim clear, it helps first to look at the historical development of ideas and controversies about time and its passage, and their consequences. To this end, I will now begin with a review of objections to the reality of time and passage that arose in classical philosophy, postponing treatment of modern objections to later chapters. It is fitting to begin with Aristotle, since, as we saw above, several of the objections just outlined have analogues in classical antiquity and can be resolved in a purely classical context; and, on the other hand, there are distinctions made by authors in the Aristotelian tradition, such as that between time and duration, which, although neglected in contemporary debates, have continuing relevance to modern issues, for instance, to the distinction between co-ordinate time and proper time of relativity theory, which we will come to in Chap. 5.

So let us take up in order the three problems mentioned by Aristotle: the reality of time, the status of the 'now', and the continuity of time.

2.3 Aristotle and the Classic Arguments for Time's Unreality

The argument for time's unreality recorded by Aristotle runs as follows:

> Some of it is past and no longer exists, and some is in the future and does not yet exist; these constitute both infinite time and the time that is with us at any moment; but it would appear to be impossible for anything which consists of nonbeings to participate in being itself. (*Physics* 217b 34-218a 3)

A more succinct version of this argument was later given by the influential medieval English philosopher William of Ockham (c. 1288–1348):

> that which is composed of non-entities is not a positive entity; but time is composed of non-entities, because it is composed of the past which does not exist now, although it did exist, and of the future, which does not yet exist; therefore time does not exist. (Ockham 1984, 496.)

But what of the present, it may be objected? Surely that exists now! Here another set of objections was already common property in ancient times. If the present has any extent, then part of it will be past and part future, so that the same argument will

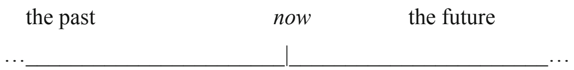


Fig. 2.1 *Now* as a boundary, not a part

apply to those parts, Aristotle argues. Only the instant dividing the past and future, the instantaneous present, can be said to exist now. But this is not properly a part of time, since it is only a boundary between past and future, binding them together into a continuous duration. (The idea is that parts must be such that adding them together will make the whole; but no matter how many instantaneous nows you add, you will not get a whole time. So the now is not a part of time, even though it is in time. See Fig. 2.1.) Thus the only parts of time are the past and the future, and neither exists.[9]

This argument for time's unreality based on the non-existence of its parts has had a long history. It was repeated by most commentators on time in the seventeen centuries between Aristotle and Ockham, including the African bishop Augustine (354-430) and the great Arab philosopher Averroës (Ibn Rushd, 1126–1198). "Time is composed of past and future," declared Averroës, "but the past has already stopped being and the future does not yet exist" (Duhem 1985, 301). And as late as the seventeenth century Gottfried Leibniz (1646–1716), Newton's main rival in natural philosophy, did not hesitate to employ a version of it against the absolute time which Newton supposed to exist independently of things. As he wrote to Newton's defender Samuel Clarke,

> Everything which exists of time and duration, being successive, perishes continually. And how can a thing exist eternally if, to speak precisely, it never exists at all? For how can a thing exist if no part of it ever exists? Nothing of time ever exists except instants, and an instant is not even a part of time. Anyone who considers these observations will easily comprehend that time can only be an ideal thing. (To Clarke, V, §49: GP VII, 402; my translation)[10]

As remarked above, modern physicists have arrived at some conclusions that sound remarkably similar. Julian Barbour and Carlo Rovelli have urged physicists to realize that time is not part of fundamental physical reality, and in Chap. 8 we will be considering these authors' more technical arguments for this based on the Wheeler-DeWitt equation, regarded as fundamental in quantum gravity. But since these physicists, especially Barbour, also appeal to classical precedents for their views, let us look at their views in relation to these precedents first.

Taking inspiration from Leibniz's championing of time as a relation among things as well as his claim in the passage above that "nothing of time ever exists except instants", Barbour writes:

---

[9] Cf. Aristotle: "time has parts, and some of them have existed, while others will exist, but none of them currently exist. The now is not a part of time… The now is a limit." (*Physics* iv, 218 a5, a23). For further discussion, see Nathanael Stein, "Aristotle on Parts of Time and Being in Time" (Stein 2016).

[10] The abbreviations in references to Leibniz's writings (here and throughout the book) are to the standard editions of his works: GP VII, 403 is page 403 in the 7th volume of Gerhardt's edition of his philosophical writings, and A VI 3, 565 is page 365 in volume 3 of series 6 of the Akademie edition of Leibniz's works. Translations are my own unless otherwise indicated.

> We have been exploring Leibniz's idea that only things exist and that the supposed framework of space and time is a derived concept, a construction from things. If it is to succeed, the only possible candidates for the fundamental 'things' from which the framework is to be constructed are configurations of the universe: Nows or 'instants of time'. They can exist in their own right: we do not have to presuppose a framework in which they are embedded. In this view, the true arena of the world is timeless and frameless—it is the collection of all possible nows. (Barbour 1999, 177)

This last claim about the "true arena of the world" being timeless is, as Barbour acknowledges, indebted to Aristotle's teacher, Plato. For in his dialogue *Timaeus* Plato had insisted on the distinction between "that which is real and has no becoming" and "that which is always becoming but never real" (27d/Cornford 1959, 16). The distinction is due to Parmenides, who claimed that if you take seriously the widely accepted axiom "Nothing can come from nothing", it follows logically that no existing thing can come to be, so that becoming must be just an appearance. Accordingly, Plato affirms, "that which is apprehensible by a rational account is what is always unchangeably real, whereas that which is the object of belief with unreasoning sensation is what becomes and passes away, but never has real being" (28a/16). Thus what is truly real is what is intelligible, what is formal, and not the particulars that appear to the senses. The influence of this Platonic doctrine down through the ages has been enormous, not least among the mathematically minded.

In a recent popular exposition of his views, Barbour recounts this Platonic doctrine of becoming as a mere appearance, and states the ambition of going one step further, to show how the appearance of becoming is produced. On his view, the world consists exclusively of what he calls "time capsules" or "instants", each of these being a concrete particular containing, in an implicit way, traces of its entire past history (Barbour dubs this conception of reality "Platonia".) These instants are "worlds unto themselves", "no thread of time joins them up." (Barbour 1999, 45ff). All we ever experience, Barbour proposes, we experience in an instant. The illusion of motion and of passage is produced by the circumstance that at any instant, "my brain contains several 'snapshots' at once", and "through the way in which it presents data to consciousness, it somehow 'plays the movie' for me". (267) Of course, there's a lot packed into that "somehow", since the very notions of presenting data and playing movies in the brain are both processes requiring not only time, but a thread linking the instants together. This is a notorious sticking point for such static views of reality: how to account for the appearance of passage or temporal becoming without presupposing the becoming of the appearance.[11]

But we needn't dwell here on the shortcomings of Platonism as a philosophy of time. It will be enough to draw attention to a fundamental ambiguity that vitiates Barbour's position. This can be done by comparing it to the views of classical authors who argued for time's unreality along the lines we were considering above. For they seem to have taken the unreality of time in a sense that is a good deal less radical

---

[11] See Shimony (1998) for criticism of this aspect of Barbour's position. See also G. J. Whitrow's criticisms in his (1961), where he quotes Hermann Lotze's apt observation: "We must either admit Becoming or else explain the becoming of and unreal appearance of Becoming." (Whitrow 1961, 311). The implicit antinomy is self-evident.

than Barbour's "timeless" interpretation. Neither Aristotle nor Averroës, for example, denied the reality of motion, nor did they believe that succession was illusory. Not even Ockham, perhaps the most relentless critic of the view that time is a kind of thing existing independently of enduring things, denied that motion occurs or that the states of persisting things exist successively. So, given the alacrity with which many modern physicists have jumped from the premise that time does not exist independently of things to the conclusion that it is eliminated from fundamental physics altogether, this is an ambiguity that we should certainly consider further.

According to Aristotle's famous definition, adopted by Averroës and Ockham, time is the number or measure of motion. As such, he claimed, it requires a soul to do the counting or measuring. From this the African bishop St. Augustine inferred that time is something existing only for the observer (Augustine 1993, Book XI, chapter XXVII)—a claim that finds echoes in some interpretations of quantum theory, as we shall see in Chap. 8.[12] But Aristotle seems to have held that the soul's counting is one thing, and what it counts another. For neither the motions it counts, such as the revolutions of the sphere containing the fixed stars, nor the succession of these revolutions, depend on being measured. On this understanding time is a kind of concomitant of motion, an aspect of the changes and motions we see about us. As Aristotle wrote, "time cannot exist without change" (*Physics* 219a 1-2). It has no independent reality, but nonetheless presupposes changes, motions and successions that do exist independently of us.

In much of his writing, Barbour speaks in exactly this vein. He speaks of the inspiration he received from the writings of the Austrian physicist, philosopher and physiological psychologist Ernst Mach (1838–1916), whom he quotes to this effect: "'It is utterly beyond our power,' he said, 'to *measure* the changes of things by *time*. Quite the contrary, time is an abstraction, at which we arrive by means of the changes of things'" (Barbour 1999, 67; Mach 1919, 224). This fits well with Barbour's masterly analysis of classical time in his (Barbour 2001), where he describes Newton as correctly perceiving that beneath the various relative times measured by the motions of the heavenly bodies, there must be an equable time by means of which they can be correlated. That is, although there is not necessarily any body performing the equable motion corresponding to Newton's absolute time, it nevertheless has a measure that "is, for all practical purposes, identical to the astronomers' ephemeris time" (Barbour 2001, 633).[13] Time in this sense is a construction with a sound empirical basis. It is an abstraction from the motions of the moon and the planets, but there is no

---

[12] According to the Copenhagen interpretation, an observable event such as the emission of a gamma ray from an atom has no independent reality prior to being brought into being by observation or measurement. Carlo Rovelli rejects the idea that measurement depends on an observer; yet he explicitly appeals to Augustine (Rovelli 2018, 180–183) in support of his view that time "is entirely in the present, in our minds, as memory and as anticipation" (182).

[13] Barbour insightfully observes that it was Ptolemy who "prepared astronomy for the day when it had to be recognized that there is no motion at all that realizes the concrete 'uniform flow of time'. … Nearly two millennia after he died, in the nineteenth and present centuries, astronomers did in fact construct an abstract time in this sort of fashion. This time, called *ephemeris time*, is the time according to which the tables and positions of the planets and moon are calculated." (Barbour 2001, 181) A similar point about Newton's absolute time being the time constructed by astronomers was

concrete motion corresponding to it.[14] Thus the sense in which it is not real is that it does not exist independently of changing things. This is the sense of the unreality of time, it would seem, that Barbour and Rovelli should be committed to. Once there is motion and change, on such a Machian view, it is not necessary also to posit time as a fundamental feature of reality. Insofar as it is something distinct from change, it is an abstraction from the changes around us.

So the theory of time maintained by Aristotle, Averroës and Ockham, and indeed Barbour in his Machian mode, is profoundly different from the radical view implicit in Barbour's Platonia. It is one thing for time to be derivative, an abstraction from changes, as it is on the Machian view, and quite another for reality to be fundamentally changeless and timeless, as it is on the view implied by Barbour and Rovelli when they talk of the "elimination of time". Where Ockham and Leibniz deny the reality of time as an entity distinct from persisting things, and construe time as depending on the supposed successions of states and the motions of such things, Barbour claims that all such successions and motions, as well as any times we might abstract from them, are simply illusions.

It's worth dwelling on this point a little longer. For one might suspect that Barbour's Platonist view seems more consistent in the following sense. How, if time is a mere abstraction, could there be things persisting through time? Conversely, if there really are things persisting through time with states succeeding one another, then surely time cannot be a mere abstraction, but must be something real? Here it is important to recognize a crucial distinction made by authors in the Aristotelian tradition that allows them to evade this dilemma, the distinction between *time,* on the one hand, and *duration* on the other. Time for them was an abstraction made from motions, in fact, a measure. But substances themselves would still endure, and still go through successions of qualities or states in the absence of any such measure being applied. A day, for example, is a measure of time derived from the rotation of the Earth on its axis (or of the Sun around the Earth, on the older geocentric view). Duration, on the other hand, is something concrete, an attribute of an enduring thing.[15] Thus Ockham subscribed to an austere universe whose only denizens are what he called *res permanentes*, persisting things. He sought to show that a statement ostensibly referring to time could be reduced to statements that made no appeal to time as an independently existing entity. Accordingly, a statement that two processes last for the same time could be parsed in terms of a coincidence between the endpoints of their durations, and a measure of the length of this time would involve comparison with the duration of a further motion taken as standard—such as that of the Sun's apparent rotation.

---

made independently in (Arthur 1995). We will return to the issue of the equability of time in Chap. 4 below.

[14] Cf. Barbour (2001, 181–2): "ephemeris time is abstract in the sense that it is not realized by any one particular motion, but it is concrete in the sense that it must be determined empirically from actually observed motions."

[15] This distinction between abstract time and concrete duration recurs, without recognition of its classical precedents, in the (early twentieth century) views of Bergson and Whitehead. See Čapek (1971) for a thorough discussion.

The same distinction between the duration of substances and time as an abstract measure occurs in Descartes' writings. He held that the duration of a thing is a mode under which we conceive it as continuing to exist, while time is a measure arrived at by comparing the durations of things to that of the motions of "the greatest and most regular motions which give rise to years and days".[16] Leibniz also upheld this distinction. Opposing Newton's identification of time with duration, he wrote to Clarke: "Everything has its own extension, its own duration, but does not have its own time and does not keep its own space." (5th Letter, §46; GP VII, 399). Duration and extension were for him attributes of enduring and extended things. Where for Newton absolute time, "without reference to anything external, flows uniformly and by another name is called duration" (Newton 1999, 408), for Leibniz time is "the order of successive things", a construction from relations among substances that are assumed to have successions of states.

Because Newton's views about space and time prevailed over those of his rivals, this distinction between time and duration has become largely forgotten. But as we will see in later chapters, there is a certain sense in which it prefigures the later distinction introduced by Minkowski between co-ordinate time and proper time. For whereas co-ordinate time is a measure of a time interval between events that varies according to the motion relative to those events of some observer measuring it, proper time is invariant, and specific to a particular process and its path. Granted, there are profound differences between these modern notions and the classical ones.[17] Nevertheless, as we shall see, the fact that there is an invariant proper time specific to any given process means that in a relativistic world we can talk about real successions and durations even in the absence of a unique time co-ordinate for the whole of spacetime.

Before we leave the topic of the reality of time, though, there is another feature of the classical argument for time's unreality from the unreality of its parts that it is worth our attention. For one may wonder how Aristotle and Averroës are entitled to hold that time is the measure of motion, given the argument they gave against time's reality. If all that exists is present, then past and future *motions* also do not exist. And if one also denies, as did Aristotle, that there can be motion in an instant, then motion and change seem just as unreal as time. Past motions do not exist, future motions do not exist, and there is no motion in the present instant.

As it turns out, we can readily resolve this conundrum, whether the argument is applied to time or to motion. This is an extremely important point: the resolution I propose here is pivotal for clarifying the issues concerning time's reality, and we will be returning to it many times in what follows. It turns on recognizing that there

---

[16] (Descartes, *Principles*, 1 §57; 1985, 212) In fact Descartes went so far as claim that the duration of a thing and its continued existence as a substance are just two ways of describing the same thing: "since a substance cannot cease to endure without also ceasing to be, the distinction between a substance and its duration is merely one of reason" (*Principles*, 1 §62; Descartes 1985, 214).

[17] In particular, the idea of a reference frame is a late nineteenth century invention absent from classical conceptions; and no classical author even dreamt of the measure of time depending on the relative velocity of a body or on its path through space and time. All this will be thoroughly discussed in Chap. 5 below.

is an ambiguity in the word "exists". When we say the past and future do not exist, we mean that they do not *exist now*. But when we refer to the existence of things in time—for instance, whether dinosaurs existed when the first humans evolved—we are talking about whether they *existed at that time*. The first landing of people on the moon does not exist in the sense of occurring now, but when cranks contend that it never really happened, it is its occurrence on July 20, 1969 that is in contention. Neither past nor future motions exist now, but if they really occurred or will occur, they did or will exist at the times of their occurrence. Past motions are those that existed during past times, putative future motions are those that will exist at future times, even though neither exist now. Thus so long as we are clear about the sense in which we are using the word 'exists', there is no confusion.[18]

The classical argument for time's unreality from the unreality of its parts itself trades on conflating these two different senses of the word exist. The crucial premise in that argument is that "only the present exists", i.e. that all and only things that are present exist. This is a very pervasive tenet, known in the trade as *presentism*. It has been subscribed to by any number of major thinkers, from Aristotle through Ockham, Hobbes and Leibniz to McTaggart and many contemporary philosophers.[19] Now it may be thought that any argument that persuades the likes of Aristotle, Ockham and Leibniz must have considerable force. But if we take the word 'exist' in the sense of 'exist now', the presentist tenet reduces to the truism: "all and only things that are present exist now", i.e. "all and only things that exist now exist now". If, on the other hand, we interpret 'exist' as 'exist at a certain time', then past events (assuming they occurred) did indeed exist when they occurred.[20] Of course, we can say that when they occurred they existed "now", namely at that previous time, and this is the sense in which presentism is true. Another way of saying this, then, is that everything that occurs indeed occurs now, but the now is the time of their occurrence, not the time at which we are talking about it. We must distinguish *the now that is the time of the event's occurrence* from *the now at which we are considering it*. To summarize: the existence of things in time is their existence at those times, not their existence now (if 'now' is understood as the time at which we are considering their existence). This is a hugely important point.

---

[18] According to Nathanael Stein, this distinction between the different senses in which things can be said to 'exist' is the gist of Aristotle's own interpretation (Stein 2016); thanks to Sean Dudley for pointing this out to me. Nevertheless, as we shall see further below, because of the limitations of language we are easily led astray on matters concerning time and existence, and we will return to the subject of the different senses of the word 'exist' in Chap. 3 below.

[19] See for example Sider (1999), Hinchliff (2000) and Markosian (2003) for modern defences of presentism, and Dorato (2006) and Savitt (2006) for criticisms. Unger and Smolin adopt a presentist position in their (2015): "By the assertion of the reality of time we mean that all that is real is real in a present moment which is one of a series of moments" (415).

[20] The same point is made by Adolf Grünbaum in his criticism of Hobbes's claim that "the present only has a being in nature", a claim which "depends on a tacit invocation of *present* occurrence as a logically necessary condition for having being or existing", so that Hobbes's claim reduces to "the mere tautology of 'only what exists now does indeed exist now'" (Grünbaum 1971, 205). For further discussion, see (Savitt 2006).

To recap the argument so far: there is more than one sense in which time may be said to be unreal. In one sense, licensed by the Aristotelian tradition, and revived by Mach and Barbour, it might mean that it does not exist independently of things, but is a kind of concomitant of or abstraction from motion. But this is very different from its being unreal in the sense that passage and change are unreal. That would have been flatly denied by Aristotle and Ockham, despite their promotion of the argument for time's unreality from the unreality of its parts. Further analysis of that argument revealed that it depends on an ambiguity in the word "exists", or its synonym, "is real": the past and future are unreal only in the sense that they do not exist now, where 'now' denotes the time at which we are considering this. They nonetheless exist or are real at the times of their occurrence if they do indeed occur at those times. (Alternatively, we can say there is an ambiguity in the word 'now', depending on whether it refers to the time of the event's occurrence or the time at which this is being considered.) The same ambiguity was seen to infect the presentist axiom that gives this argument its apparent cogency, namely the premise that all and only things that exist now are real. If 'now' is taken as the time of these events' occurrence, then it reduces to a truism, whereas if 'now' is taken as the time at which the statement is uttered, it is merely false.

Still, one might insist, when we say past motions existed at past times, or that future ones will exist at future times, doesn't this presuppose the existence of the past and future times at which those motions must exist? If so, this becomes another motivation for adopting a static view in which the now is excluded. If past events exist at past times, and future ones at future times, then all events alike have been presupposed to exist, and there is no place for temporal becoming. In this form, the static view is usually called *eternalism*. It is the view that all events and their temporal (and spatiotemporal) relations exist eternally, so that their coming into existence is precluded.

There is certainly an innocuous sense in which all the events and temporal relations in spacetime can be said to exist or to be real. 'Real' is a contrast word, and the contrast here would be with a view that denied the existence of the material universe and the processes in it. But we must be very wary of using 'exist' in a temporal sense here, as advocates of eternalism do in their less guarded moments when they talk of future events being "already" real, and so not needing to become. Leibniz had the right response to the question of the existence of times when he wrote in an early dialogue: "Time itself ought not to be said to exist or not to exist at some time, otherwise time would be needed for time." (A VI 3, 565; Leibniz 2001, 209). He made a similar point in his criticism of Samuel Clarke in the last year of his life, when he objected to Newton's ally's describing duration as "immutable and eternal" (Clarke, 4th Letter, §10; GP VII, 383). "One cannot say that a certain duration is eternal," wrote Leibniz, "but one can say that things which endure always are eternal, in gaining always a new duration." (To Clarke, V, §49: GP VII, 402). Just as a time cannot be said to exist at a time, so a duration cannot be said to exist at all times. Duration and time are not existents in the sense that they exist at times. But this does not make them illusory.

Now, one might claim that *temporal relations* are eternal in the sense of existing timelessly: that the fact that some event occurs *before* another (say, Plato's being born before Aristotle) is not something that changes in time. This would be in keeping with philosophical and theological tradition, where things that do not exist in time—numbers, for example, or God—have been called eternal. (In the same way, Plato had claimed that what are real are the eternal forms, whereas what comes to be in time is mere appearance, as discussed above.) Such claims, however, are by no means unproblematic. God, for example, is supposed to act, and a timeless being can hardly be said to act.[21] This is perhaps why Newton, in defiance of Christian theological tradition, identified God's eternity with his existence through infinite time. This means that God, instead of existing outside time, exists at all times—a notion that had traditionally been called sempiternity (from the Latin for 'always', *semper*), in distinction from eternity.[22] This was regarded as theologically suspect, since it would make God an actor in his own creation as opposed to transcending it, starting a slippery road down to a view like Spinoza's, where God is simply another word for Nature. But whatever difficulties may attend these traditional conceptions, it is not helpful to regard time and temporal relations as eternal, if only because the eternal—that which exists beyond time, and thus at no time—tends to get confused with the sempiternal or everlasting—that which exists at all times.[23] Finally, even if we regard *temporal relations* as existing eternally (timelessly), we cannot so regard *events*. *Events exist neither at no time nor at all times, but at the time of their occurrence*. So past events or motions cannot be said "already" to exist on the grounds that they exist at *all* times: they exist only when they occur. Eternalism, it seems, is no more viable an alternative than presentism.

Thus the first major source of difficulties about time can be seen to lie in a lack of clarity concerning the senses of existence and reality. This afflicts the traditional argument for time's unreality from the unreality of its parts. Like its *presentist* premise, that argument was seen to be vitiated by a failure to distinguish between 'exists' or 'is real' in the sense of really occurring at the time of its occurrence, and 'exists now', i.e. exists at the time of that utterance. The rival *eternalist* view in a sense commits

---

[21] This is a knotty problem in theology. In orthodox Catholicism, God exists outside time, but can choose to be immanent within it—as, crucially, He is, in the form of Christ. See (Kneale 1960–61) and (Sorabji 1983, 136). Certainly, it was extremely important for Newton that God could be said to exist in time, to intervene in the world's operations. This was one of the main points of contention in his and Clarke's controversy with Leibniz, who took the view that (setting aside miracles) God would act through the actions of his creatures, in keeping with his omniscient anticipation of all their actions.

[22] The distinction between eternity and sempiternity is most closely associated with Boethius, *De Trinitate*, 4, ll. 64–77; see Kneale (1968–69) and (Sorabji 1983, 116).

[23] Even Leibniz is guilty of this confusion, when he magnanimously allows in his criticism of Newton and Clarke (Fifth Paper, §49; G VII, 402) that "if by saying that the duration of a thing is eternal it is only meant that the thing endures eternally, I have nothing to say against it", and that a thing's enduring eternally is to be understood as its "gaining always a new duration" (my translations). But on Leibniz's own conception of eternity, God is eternal only in the timeless sense. "Gaining a new duration" is in keeping with Newton's equating of God's eternity with sempiternity, but not with timelessness.

the opposite mistake, that of taking events and their temporal (and spatiotemporal) relations to exist eternally in the sense of *existing at all times*. But time and temporal relations should not be said to exist in time, while events exist at the time of their occurrence, not at all times. Thus if we keep firmly in mind what kind of existence we are attributing to things we can avoid these snares, and evade this alleged dichotomy between eternalism and presentism altogether.[24]

2.4 Motion and the Flow of Time

A second major source of difficulties concerning the flow of time is that it seems to involve events' somehow changing their status with respect to the *now*. Those events which yesterday we regarded as happening now, we do not regard as now today. As Aristotle indicated, there are two ways to describe this state of affairs, but both seem to lead to paradox. One way is to admit that what is now is always different from one time to another. But by itself this does not seem enough, as we also want to incorporate the idea that time *passes* from the earlier to the later. Many interpret this passage of time as requiring that the *now* itself move along the time axis from an earlier to a later time. This, as we shall see, does not appear to be a coherent notion. A motion of the *now* along the time axis necessarily presupposes a second time dimension with respect to which its motion would be calibrated. But the other alternative offered by Aristotle is equally problematic, and this is to regard the now as "always the same thing". This suggests a presentist scenario in which "we" remain rooted in the now, and future events come gradually closer until they come to be now (and are therefore experienced by us), and then recede ever further into the past. But this also seems incoherent. For whether it is the events that are fixed and the now that is moving, or the now that is fixed while future events move up through it and into the past, such a relative motion would presuppose another time dimension, and thus lead to an infinite regress of times.

The unacceptability of either of these alternatives was exploited by the idealist philosopher J. M. E. McTaggart in the early twentieth century in his argument for the unreality of time—an argument that has been highly influential in subsequent thinking about time. I maintain that it is premised on the same kind of mistake as we have just analysed in connection with presentism vs. eternalism, namely the idea that events pre-exist in a temporal ordering, a temporal ordering that is prior to or independent of their coming to be. We'll return to his argument in Chap. 3 below. In the meantime, though, it is worth staying with Aristotle, and his own solution to the problem of the status of the 'now'.

For Aristotle, time is closely connected to motion. As we saw, he defines it as "the measure of motion". This is not a claim that either Newton or Leibniz saw fit to contest, despite the very different conceptions they had of time's nature, and

---

[24] I return to this issue in Chap. 3 in connection with the "block universe" view, which is often equated with eternalism, and also in Chap. 6, in the context of the present in relativity theory.

through them it became part of the bedrock of modern physics. (Indeed, as we shall see in Chap. 5, the close connection between time and the measure of motion was seminal for Albert Einstein, who exploited it in constructing his Special Theory of Relativity.) Because of this mutual involvement of the measure of motion and time, it is natural to relate the issue of the changing now to that of the change of position involved in motion. Thus in his *Physics* Aristotle writes:

> A *now* follows a moving object, just as time follows change; for it is the moving object that enables us to know before and after in change… So time is not only continuous thanks to the now, but is also divided at the now, because this too follows the nature of the movement and the moving object. (Aristotle, *Physics*, IV, 219 b26-28, 220 a4-6).

By the *now* "following a moving object", Aristotle means that we can tell when one event is before another when we follow a motion of the first to the second (think of following the moving hand of an analogue clock or stopwatch). Here Aristotle apparently takes his cue from a mathematical tradition that dates back to the Pythagorean thinker, Archytas of Tarentum. The latter was famous for having solved certain problems in physics and mathematics with his kinematic conception of curves. A spiral, for example, can be conceived as the curve generated by a point moving with a constant speed along a straight line that is rotating (also at a constant speed) about a centre. In this tradition a line would be generated by the movement of a point. Correlatively, it was thought, time would be generated by the movement of the *now*—a conception upheld by Newton, as we shall see. But the idea that a self-identical now is moving through time is problematic, as will be argued in more detail in Chap. 3 below.

Aristotle was more circumspect. Appealing to the close relationship between time and motion whereby each measures the other, he notes that "what is before and after is found primarily in place", and by analogy is therefore found in change. "And since time always follows the nature of change, what is before and after applies also to time." Time is not to be identified with change, but it is "that feature of change that makes number applicable to it." So 'now' is "single and identical" in the same sense that the number 10 is the same whether we count ten chickens or ten foxes. But the various successive nows at which a moving thing is found in successive places are different, just as the foxes and chickens are different things numbered, even though the two collections are both ten in number.

Aristotle's point can be illustrated by reference to the drawing of a line from left to right with a sharp pencil. As the line is drawn, the point of the pencil passes through all the points that will be on the line when it is completed. Each point corresponds to an instant of the time (or 'now') during which the line was drawn. But if we thus identify the nows with the instants, then in this sense, as Aristotle says, "the nows are always different". So there is no motion of a now in the sense of a self-identical entity moving along a time-axis. There is, however, passage, the tracing of the line by the motion of the tip of the pencil from left to right. But if there is passage from one spatial end of a motion to the other during a given interval of time—a spatial before and after in motion—why should there not also be transition from one temporal end

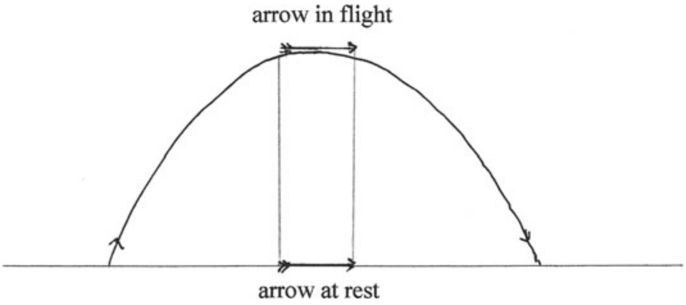


Fig. 2.2 Zeno's arrow paradox

of the motion to the other as the space is traversed? The before and after in time would "follow" the before and after in changes of place, to use Aristotle's way of speaking. If there is no lapse of time then the pencil tip cannot trace the line; if it moves from left to right across the page, any point on the left was drawn earlier than one to its right. In this sense, the reality of passage stands or falls with the reality of motion.

As Aristotle well knew, however, the reality of passage had been contested on these very grounds by Zeno of Elea, who denied the reality of motion. So let us turn to a consideration of Zeno's paradoxes, and Aristotle's responses to them.

In his dialogue *Parmenides*, Plato reports that Zeno devised his paradoxes to pay back in kind the critics of his teacher and lover Parmenides, who held that reality was changeless, and motion but a deceptive appearance. According to Aristotle, in the third of his paradoxes designed to refute the reality of motion, Zeno claimed that

> If everything that occupies a place equal to itself is at rest, and if a moving object is always occupying such a place in the now, then a moving arrow is motionless. (*Physics* VI, 9: 239b 5-8)

The idea is that at an instant the moving arrow in a certain place would occupy exactly the same place as one at rest at that instant. This is now known as Zeno's Arrow Paradox (Fig. 2.2).

In a more modern idiom, it might be re-expressed as follows: Granting that there cannot be motion in an instant, it follows that the arrow is not moving at any instant or *now* of its flight; but if it is not moving at any single instant while it is in flight, then it does not move at all. Aristotle dismisses this paradox as a fallacy of composition: "Here the conclusion depends on assuming that time is composed of *nows*; if this assumption is not granted, the argument fails." (*Physics* VI, 9: 239 b29-32). A time can no more be composed of instants than a line can be composed of points, or a number from zeros. Elsewhere in the same chapter Aristotle says more. An indivisible for him is an endpoint of a magnitude or interval. Thus a point is the indivisible beginning or endpoint of a line, and a *now* is the indivisible beginning or endpoint or beginning of a stretch of time. There can be motion across a stretch, but not at an instant or now. Thus for Aristotle "nothing moves in the now" (*Physics* 234a 24);

but by the same token, “nothing can be at rest in the now” either (*Physics* 234b 7), since motion and rest both require an interval of time. (To be at rest something must remain in the same place, and *remaining* requires time.) So he can cede to Zeno that the arrow is not moving at any and every now without at all conceding that it is motionless.

There is nothing incoherent about this response. But it does have one very curious consequence, which Aristotle explicitly recognizes. This is that even though at any instant of its motion Zeno’s arrow is not moving, it will nevertheless be true to say at any instant of the motion that the moving arrow *has moved*. This consequence is counterintuitive, but it is not absurd. In fact, as we shall see in a moment, a widely accepted modern theory of change has exactly the same consequence. This is the so-called “at-at” theory of motion advocated by Bertrand Russell in 1903, and still accepted by many contemporary philosophers of science.

But before we address that, let’s consider a more natural modern response to Zeno’s Arrow Paradox. This is to appeal to the concept of instantaneous velocity, made precise by Newton’s and Leibniz’s invention of the differential calculus in the seventeenth century. Nowadays we are familiar enough with the concept: instantaneous velocity is what is indicated on the speedometer of a car (or on the reading of a policeman’s radar gun!). The basic idea is that it is the velocity with which the arrow (or speeding car) would traverse some small finite distance in the finite interval of time immediately afterwards, assuming it neither accelerated nor decelerated. Mathematically, it is the limit of a sequence of ratios of distance travelled to time taken, as the time is shrunk to the instant marking the beginning of the interval. For such an object as the moving arrow or the car, this limiting ratio is itself finite, but not zero. So even though strictly speaking there is no motion in an instant, you can still define a finite velocity at an instant. A moving arrow has a finite, non-zero velocity; an arrow at rest has zero velocity.

Rather surprisingly, however, Bertrand Russell summarily dismissed any such attempt to resolve Zeno’s Arrow Paradox by appealing to the calculus, and took Zeno’s side. Expressing himself with characteristic brio, he vehemently rejected the idea of the moving arrow’s having a state of motion or change at each instant:

> People used to think that when a thing changes, it must be in a state of change, and that when a thing moves, it is in a state of motion. This is now known to be a mistake. When a body moves, all that can be said is that it is in one place at one time and in another at another. (Russell 1929, 83-84)

In his earlier *Principles of Mathematics* Russell had written in the same vein:

> “If everything is in rest or in motion in a space equal to itself, and if what moves is always in the instant, the arrow in flight is immovable.” This has usually been thought so monstrous a paradox as scarcely to deserve serious discussion. To my mind, I must confess, it seems a very plain statement of a very elementary fact, and its neglect has, I think, caused the quagmire in which the philosophy of change has long been immersed. In Part VII, I shall set forth a theory of change which may be called *static*, since it allows the justice of Zeno’s remark. (Russell 1903, 350)

In Part VII of that work, Russell outlines the theory of motion and change in question, now known under the moniker the *at-at theory of motion*. According to this theory, all there is to motion is an occupation of different places at different times: “when different times, throughout any period however short, are correlated with different places, there is motion; when different times, throughout any period however short, are all correlated with the same place, there is rest.” (Russell 1903, 473). And motion consists “*merely* in the occupation of different places at different times, subject to continuity … There is no transition from place to place” (473). Russell’s denial here of any transition or passage from place to place explains why he calls his analysis a *static* theory of change.

Russell’s at-at theory has been highly influential, especially in the form in which it was promulgated by Wesley Salmon,[25] and its adoption by philosophers such as Adolf Grünbaum and Jack Smart has done much to cement the “static view” as the dominant philosophy of motion and change among scientifically minded philosophers. There are, however, some troubling ambiguities in Russell’s own statement of the theory, and these ambiguities, though subtly transformed, continue to bedevil contemporary statements of the static view. For example, in a further passage from his (1903) reprinted in his later *Mysticism and Logic*, Russell says

> Weierstrass, by strictly banishing all infinitesimals, has at last shown that we live in an unchanging world, and that the arrow, at every moment of its flight, is truly at rest. The only point where Zeno probably erred was in inferring (if he did infer) that, because there is no change, therefore the world must be in the same state at one time as at another. (Russell 1903, 347; 1929, 80-81)

We’ll come back to Weierstrass in a moment. For now the important thing to note is that in this passage Russell explicitly denies change and motion, whereas in the description of the theory in Part VII of the *Principles* he is only denying an instantaneous state of motion, not motion and change themselves. Here he claims the arrow is at rest at every moment of its flight, whereas according to the analysis in Part VII the arrow would only be at rest if at the different times of its flight it were at the same place.

The more charitable interpretation of Russell’s theory of change, therefore, is to understand it as reducing change to *temporal difference*, while denying that there is any “transition” from one state or event to another. This seems to be Russell’s meaning when he remarks that

> change is due, ultimately, to the fact that many terms have relations to some parts of time that they do not have to others. But every term is eternal, timeless, and immutable; the relations it may have to parts of time are equally immutable. It is merely the fact that different terms are related to different times that makes the difference between what exists at one time and what exists at another. (Russell 1903, 471).

From this point of view there is motion from point A to point B if something is at A at an earlier time, and at B at a later time. Thus there is no necessity to deny motion, as in Russell’s reckless endorsement of Zeno. On the more sober interpretation of

---

[25] Cf. (Salmon 1984, 110): “To get from point A to point B consists merely of being at the intervening points of space at the corresponding moments of time. There is no further question of how the moving arrow gets from one point to another.”

his at-at theory—the official version he gives in *Principles,* Part VII—the arrow can only be said to be at rest or in motion *across an interval of time*, for only then can the correspondence demanded by the at-at theory be established between the places it is at and the times it is at them. In this respect, the sober version of the at-at theory is in complete conformity with Aristotle's position discussed above. It is no more true on this view that the arrow is at rest at each instant than it was on Aristotle's. And to infer the unreality of motion from the fact that there is no motion at each instant is the very fallacy of composition pointed out by Aristotle, just as invalid on the sober at-at view as it was on his. It will even be true on the at-at theory, as it was on Aristotle's, that whilst the arrow can never be said to be *moving now*, it can be said to *have moved* at any instant of its motion.

So our preliminary Aristotelian conclusion stands: the reality of passage stands or falls with the reality of motion.

2.5 The Continuity of Time

But why did Russell want to deny that a moving body is in a state of change at each instant, and what moved Grünbaum and Smart to follow his lead and deny the reality of passage? A good part of their motivation derives from difficulties they find in representing the continuity of change mathematically—the third of the problems concerning time that Aristotle had identified. An examination of these issues will give us some further insight into the origins of the static view.

There are two main factors here. One of them was Russell's understanding of the mathematical notion of continuity as irreconcilable with the idea of a state of change. As we shall see, this depended on his idiosyncratic notion that a state of change involves actual infinitesimals, which he regarded as precluded by the new foundations of mathematics laid down by Weierstrass and Cantor. But the more important factor was a consideration already noted by Aristotle, and recognized by him as underlying Zeno's other paradoxes intended to show the impossibility of motion or change. This is that a continuous line is such that between any two points there lies another point—a property that we call *denseness*. This property of the continuum entails that to any point in it *there is no next point*, and likewise, that to any instantaneous state in a continuous change, *there will be no next state*. But how can something come to be out of a previous state if there is no immediately preceding state?

Again, the property of denseness seems to rule out any beginning of motion, a fact already exploited by Zeno in his Dichotomy Paradox. For in order to move a certain distance, Zeno's Arrow would first have to traverse a half of this distance; and before it could get to the midway point, it would have had to have got a quarter of the way there; and so on to infinity. But no matter how close a point is to the starting point, there is always a further point between the two it would already have had to have reached first. So the motion could never have begun.

This difficulty was lucidly explained by Alfred North Whitehead, Russell's friend and collaborator on the monumental work on modern logic, *Principia Mathematica*, and author of a rival theory of gravity to Einstein's that rejected the idea of a curved spacetime. In his *Science and the Modern World* (1925), Whitehead argues that the denseness of moments in the continuum rules out their being consecutive.[26] In fact, anyone giving this objection based on denseness

> assumes that Zeno understated his argument. He should have urged it against the current notion of time itself, and not against motion, which involves relations between time and space. For, what becomes has duration. But no duration can become until a smaller duration (part of the former) has antecedently come into being… The same argument applies to this smaller duration, and so on. Also the infinite regress of these durations converges to nothing—and even to the Aristotelian view that there is no first moment. Accordingly time would be an irrational notion. (Whitehead 1925, 127)

Whitehead himself proposed to avoid this difficulty by advancing an atomistic theory in which becoming occurs in discrete units, and the continuous extension of time is a retroactive construction in experience (Whitehead 1925, 127). "Continuity," he insisted, "concerns what is potential; whereas actuality is incurably atomic" (Whitehead 1930, 95). For him "every act of becoming must have an immediate successor, if we admit that something becomes. For otherwise we cannot point out what creature becomes as we enter upon the second in question…" (107). Consequently he proposed that the creatures which become are "actual occasions" that come into being discretely, where each "actual entity is an act of experience" (105). We'll return to the issue of whether becoming is only constituted in experience in Chap. 3. But the idea of avoiding Zeno's paradoxes by asserting the atomicity of becoming has an ancient pedigree, going all the way back to Aristotle's younger contemporary, Diodorus Cronus (d. 284 BCE).

Diodorus accepted Aristotle's conclusion that something could have moved without it ever being true that it is moving now. But he interpreted this as a movement through a partless place in a partless time, or time atom. Taking Aristotle's characterisation of the now as an indivisible, he construed time as composed of nows, although for him these were not instants, but discrete atoms of a positive finite size.[27] Contemporary with him was the Greek atomist Epicurus (341-270), inspirer of the Roman poet Lucretius (*c.* 95–55) (and modern lovers of food!). According to Demetrius of Lacon—an atomist follower of Epicurus from the second century BCE—Epicurus taught that each thing emerges from a previous one in a next, minimal time. At around the same time in northern India a version of atomism was developed, probably independently, by the Buddhist sect of the Sautrãntikas of the second or first century BCE. They proposed a world of "point-atoms" or durationless events (*dharmas*), each ceasing to exist as soon as it comes into being. As we'll see when we come to

---

[26] Whitehead notes that a similar argument for the denseness of time can be found in Kant's Axioms of Intuition in his *Critique of Pure Reason*. But, he observes, it contradicts an earlier claim Kant makes in the same section that in the smallest portion of time one can only think "the successive progress from one moment to another" (Whitehead 1925, 125–6).

[27] Here I am following the interpretation of Sorabji (1983, 16–21).

consider their views again in Chap. 6, there is nothing in their fragmentary world corresponding to Aristotle's *now* (το νυν), a plane of becoming that is experienced by everyone existing at the same time, and so no discrete world state, as is implicit in the time atomism of Epicurus and Diodorus. We find perhaps a combination of these views in the *kalam* atomism of the Arabic-Muslim school in the 9th C of the common era, made known in Europe by Maimonides. These thinkers posited a world that flashed in and out of existence at each time atom, as did, later still, the Occasionalist followers of Descartes in the seventeenth century.[28] But the difficulty with all such atomist theories of space and time has always been to reconcile them with the continuity that seems necessary for geometry. As mentioned already, however, modern theory is revisiting this idea. In the search for a theory of quantum gravity theorists have found it necessary to propose that (on an unimaginably small scale) spacetime might in fact be granular after all.

But does the objection that the instants of time are dense constitute an obstacle to the reality of motion? Aristotle did not think so. This is because he rejected a premise he saw as underlying Zeno's argument, namely the assumption that a line had to be composed of the points you could identify in it, or that a continuous time must be composed from the instants that could be assigned in it. Rather, according to Aristotle, points and instants only designate locations at which the continuity *could* be broken: a point marks where a continuous line could be divided, an instant where a continuous time could be divided. There are infinitely many points or instants where they *could* be divided, but so long as the line or time is not actually divided, there are no actual points or instants in it. Zeno's Dichotomy argument is thus circumvented: there is no actual instant between any point of a continuous motion and its beginning, although there is a potential infinity of possible instants, denoting points at which the line could be divided. Similarly, there is no first instant after the beginning of a continuous duration that must be passed before it can get to any subsequent one.

Modern authors have generally not found this response adequate, just because of the account of the continuum given by Weierstrass, Dedekind and Cantor, to which Russell had alluded. For according to Cantor's transfinite set theory, there are actually infinitely many points in a given line—not just potentially infinitely many, as Aristotle had contended. Moreover, there are more than can be counted. That is, if you try to set them in a 1-1 correspondence with the natural numbers—1, 2, 3, 4, …—there will always be some you have missed, as Cantor successfully proved. Now, you might think that this is evident, given the denseness property: between any two points there are always more. What was surprising about Cantor's result was that he had already proved that denseness alone is insufficient to constitute continuity. For if you assign a number to every place where the continuum can be divided into parts of a given ratio—1/2 for dividing it in two, 1/237 for dividing it so its parts are in the ratio of 236:1, and so forth—each of these points of division will correspond to a rational number. But, Cantor proved, the rational numbers *can* be set in 1-1 correspondence with the natural numbers, and are therefore countable. Yet this will

---

[28] For a discussion of these atomist views, see (Sorabji 1983, Chaps. 2 and 24), and also (Arthur 2012).

not include points corresponding to irrational numbers—for instance, if the line is divided into parts having the ratio $1{:}\sqrt{2}$. If we include irrationals along with rational numbers, we have the real numbers, and these will correspond to all the points in a continuous line, the "real line", as established by Weierstrass. But Cantor showed that the real numbers cannot be set in 1-1 correspondence with the natural numbers: there is an uncountable infinity of them (a non-denumerably infinity, to use the technical jargon) in any segment of the real line. Thus in Cantor's set theory a line is regarded as an infinite aggregate of points, but of which there are non-denumerably infinitely many. Among the significant results of this theory was its application to measure theory. For here it can be shown (as stressed by Grünbaum) that even though the points all have measure zero, and a countable number of them still has measure zero, a non-denumerably infinite set of them can be proved to have a finite measure. So the incompletability of the divisions in the continuum does not oblige us to follow Aristotle and deny that the points marking divisions in it are actual. The points, according to Cantor, are actual, there are infinitely many, and yet (in the sense of measure theory) a non-denumerably infinite set of them can still constitute a finite measure.

In his *Principles of Mathematics* Russell took this account to show that the older way of conceiving continuous quantities in mathematics was false. One such account that he took as representative was that given by the neo-Kantian philosopher Hermann Cohen, whose *Das Prinzip der Infinitesimalmethode* (1883) had become the blueprint for the influential Marburg School of neo-Kantianism.[29] Without going into details, the brunt of Russell's disapproval fell on Cohen's idea that the reality of extensive magnitudes such as space, time and velocity, derives from their being generated from infinitesimals interpreted as *intensive magnitudes*.[30] For instance, a velocity v is generated from its infinitesimal dv, and a time t from the infinitesimal moment, dt. In fact, though, this was an idiosyncratic reading of the calculus. Not even Cohen's allies and students could endorse his view that a derivative such as ds/dt presupposed the reality of ds and dt as actual infinitesimal (albeit, intensive) magnitudes. As was recognized by Natorp and Cassirer, both graduates of the Marburg School, all that is necessary to represent a state of change, such as an instantaneous velocity, is for its distance traversed s to be represented as a function of the time t, and for there to exist a limit of a sequence of finite ratio of increments of distance to increments of time.[31]

---

[29] See Monier-Williams (2007). It was in opposition to Cohen's neo-Kantianism that both the logical positivism of Rudolf Carnap and the phenomenology of Martin Heidegger were conceived: see Friedman (2000).

[30] The distinction between an extensive magnitude—one that can be represented by an extended line—and intensive magnitude (which cannot) goes back to the doctrine of the *intension and remission of forms* of the Oxford Calculators. According to Kant, "In all appearances, the real, which is an object of the sensation, has intensive magnitude" (B207), "intensive magnitude" being "a measure of how an object 'fills' (*erfüllt*) space or time." (Jankowiak 2013, 387).

[31] Thus, if in a given finite interval of time Δt after some time t, a moving body has traversed a finite increment of distance Δs (the difference $s(t + \Delta t) - s(t)$), then the instantaneous velocity ds/dt represents the limit of a sequence of finite ratios $\Delta s/\Delta t$ as Δt is made to become arbitrarily small.

Russell, newly aware that Weierstrass had given a successful definition of continuity that made no reference to actual infinitesimals in the continuum, and accepting Cantor's claim to have refuted them, rejected Cohen's way of conceiving continuity as untenable. He saw Cohen's locating of the generation of quantities by infinitesimals as "an attempt to extend to the *values* of a variable the variability which belongs to it alone" (Russell 1903, 351). "People," he explained, "picture a variable to themselves as successively assuming a series of values, as might happen in a dynamical problem" (344). But on the "static theory of the variable" provided by Weierstrass, all that exist are values of position at different times, the difference between which is always finite. From this Russell inferred that Zeno was correct in denying the very idea of a state of change. For even though Russell accepted the definition of derivatives through limiting processes, he nonetheless denied that velocity and acceleration are physical facts. They are not "properties belonging *at each instant* to a moving point", he contended, but "merely real numbers expressing limits of certain ratios". These numbers are values of a variable, and it is a mistake to interpret them as varying over time.

But here Russell has over-interpreted his critique of Cohen. The sense in which the values of a variable are "static" is a mathematical one, not a metaphysical or physical one, involving a lack of motion or change. One does not have to be committed to the existence of infinitesimals in the continuum to believe that a derivative—such as ds/dt—is not only mathematically well-formed, but has a correlate in physical reality, contrary to Russell's claims. The state of change depends only on the existence of time-derivatives, and does not stand or fall with the existence of infinitesimals in the continuum.[32]

The valid core of Russell's criticism is this. If one assumes (as Cohen appears to)[33] that Zeno's arrow progresses from point x to its next position at point $x + dx$ by the accession of the infinitesimal element dx, then this is incompatible with the *denseness* property of the continuum mentioned above, which entails that there is "no consecutive moment or next position" (Russell 1903, 473). But this seems contrary to how things appear to the senses, where any given event is perceived as coming into being out of the preceding one. There is thus an incompatibility between the sensed consecutiveness of coming into being and the denseness of the mathematical continuum, just as Whitehead had observed.

---

[32] As I have argued elsewhere, this is a symptom of a deeper difficulty with the at-at theory. This is its founding assumption that "motion consists *merely* in the occupation of different places at different times, subject to continuity", that is, the assumption that the existence of a one-one correspondence between the point-events in a continuous process and the instants in a temporal continuum is sufficient for motion. On the contrary, this is insufficient structure with which to define motion, which also requires a tangent space at each spacetime point, and a temporal orientation for the spacetime.

[33] Russell notes that Cohen characterizes infinitesimals as intensive magnitudes, but he objects that intensive magnitude should not be confused with an infinitesimal extensive magnitude: "for the latter must always be smaller than finite extensive magnitudes, and must therefore be of the same kind with them; while intensive magnitudes seem never in any sense smaller than any extensive magnitudes" (Russell 1903, 344).

Subsequently, under Whitehead's influence, Russell tried to resolve this difficulty in his *Our Knowledge of the External World* of 1914 by giving a construction of mathematical time from the sense data of experience.[34] Reasoning that instants are not among the data of experience (1929, 123–4), he used Whitehead's method of extensive abstraction to construct them out of events of extended duration. This was not a resounding success, however. For in order to get his construction to deliver the required denseness of mathematical time, Russell was forced to make some empirically dubious postulates about the overlapping of sensed events.

In response to these difficulties Adolf Grünbaum proposed that it is a mistake to try to derive the temporal precedence of instants from the relation of succession given in intuition: "If the 'later-than' relation is defined intuitively, it is clear that no dense temporal order of events can be created by an ordering relation whose very meaning involves 'nextness' and hence excludes denseness" (1950, 165–166). He contended that this difficulty could be avoided by adopting an ontology of *point-events*—by "postulating point-instant-particles, i.e. *events*, rather than things, to be the basic entities of nature" (152). These point-events, unlike sensed events, could be assumed to exist in one-one correspondence with the points in a Cantorian continuum.[35] Such events, he claimed, "simply are or occur (irreversibly in some cases) but they do not 'advance' into a pre-existing frame called 'time'." (172). Similarly, Jack Smart insisted that "Events do not come into existence, they occur or happen. 'To happen' is not at all equivalent to 'to come into existence' and we shall be led far astray if we use the two expressions as though they could be substituted for one another." (Smart 1949, 486). Thus, these authors are arguing, since there is no coming into existence of point-events out of immediately preceding ones, and all the point-events in a continuous process can be put into one-one correspondence with the points in a Cantorian continuum, there is no becoming or passage in any process.[36]

But in fact this is a *non sequitur*, and—for all the sophistication of Grünbaum's appeal to measure theory and non-denumerable infinities of points and instants—it is the same one that we already exposed in the previous section when discussing motion. It is a fallacy to infer a lack of passage across an interval from a lack of

---

[34] In the Preface to the first edition of this work, dated June 1914, Russell ascribes almost all the changes in his views from his *Problems of Philosophy* (London and New York, 1912) to the influence of Whitehead, in particular "the suggestion for the treatment of instants and 'thing', and for the whole conception of the world of physics as a *construction* rather than an *inference*" (Russell 1929, viii).

[35] "Clearly", Grünbaum wrote, "these events are not sensed and their properties differ fundamentally from such perceived 'events' as the sensed coincidence of thunder and lightning. … *the events constitute a linear Cantorean continuum with respect to the relation 'later-than'! In this way the concept of 'later-than' becomes the key to the temporal order without involving the nextness property*." (1950, 152, 168–9)

[36] "*Since individual events do not move*, but are nevertheless ordered non-consecutively by 'later-than', it is the theory of events as here interpreted which justifies Russell's writing: 'Weierstrass, by strictly banishing all infinitesimals, has at last shown that we live in an unchanging world…'". (Grünbaum 1950, 179)

passage at any instant. Russell and Grünbaum both grant that *motion* occurs across an interval, even though (if the points in this interval are dense) there is *no next point* and so no movement from one point to the next. So there is no reason *passage* cannot occur through an interval despite the denseness of point-events. Even if there can be no becoming of one such event out of its immediate predecessor, as Whitehead had required of becoming, this does not entail that such a point-event cannot have become.

Now, there is no doubt that this grates with our intuitions. But this is an exact analogue of what was said above in connection with both Aristotle and the at-at theory: at any *now*, the arrow can be said to have moved, even if on those theories it cannot be said that it is moving at each *now*. Graham Priest has articulated the unease that many will feel:

> … it does not ease the discomfort … when one tries to understand how the arrow actually achieves its motion. At any point in its motion it advances not at all. Yet in some apparently magical way, in a collection of these it advances. Now a sum of nothings, even infinitely many nothings, is nothing. So how does it do it? (Priest 2006, 218-9)

The fact is, however, that it is not a sum of nothings: that was exactly Aristotle's point in denying that time is a collection of instants. Time is not composed from instants, nor is motion from instantaneous states. Even the supposition of a non-denumerable infinity of instantaneous nows does not alter the fact that between any two there are always more (denseness), so that there is no next now and no first now after the beginning of a duration. But provided a point-event is the terminus of a continuous process, there is a passage from earlier events to it, and it can therefore be said to be the terminus of a process of becoming. If you deny passage from one point-event to another, then by parallel reasoning you are obliged to deny the reality of motion, the passage from one point to another.

So none of the third set of classical objections to passage relating to its absence in the now has hit its mark. Russell's attempt to infer a static theory of change from the Weierstrass-Dedekind-Cantor mathematical model of the continuum was seen to involve several errors. The values of a variable are not constant in the sense that they remain unchanging through time, so that it is a category mistake to enlist Weierstrass' static theory of the variable in support of the idea that reality is unchanging. Secondly, Russell's assumption that the idea of a state of change stands or falls with infinitesimals cannot be accepted, since the derivatives characterizing an instantaneous state of change are regarded as perfectly well defined without need of an appeal to infinitesimals. Most importantly, though, Russell's and Grünbaum's inferences from the at-at theory to the impossibility of passage are themselves invalid: from the fact there is only a point-event or instantaneous state of change (and no actual becoming) at each instant in an interval, it does not follow that there is no passage in that interval, any more than it follows from the fact that there is only an instantaneous velocity (and no actual motion) at each instant that there is no motion over the interval. If the arrow has moved, then it has passed through both space and time. A change has occurred over an interval of time; but that is impossible unless there is passage, since passage just is change in time.

With this survey of the classical objections to the reality of time and passage behind us, we are now in a good position to tackle contemporary objections to time flow, which depend on many of the same misconceptions.

2.6 Summary

- In this chapter I have argued that claims that time is unreal are ambiguous between two senses: time can rightly be regarded as a *derivative* concept, an abstraction from motion and change that does not need to be posited as something existing absolutely; but this does not license the claim that it is *eliminable*, that passage and change are unreal. That would have been flatly denied by the likes of Aristotle and Ockham, who presuppose things as enduring and changing, with their changes really succeeding one another.
- The classical argument for the unreality of time from the non-existence of the past and future was found to turn on an ambiguity in the word "exists", or its synonym, "is real": the past and future are unreal only in the sense that they do not exist *now*, where 'now' denotes the time at which we are considering this.
- The same ambiguity was seen to infect the *presentist* axiom that gives this argument its apparent cogency, namely the premise that all and only things that exist now are real. If 'now' is taken as the time of these events' occurrence, then it reduces to a truism, whereas if 'now' is taken as the time at which the statement is uttered, it is merely false.
- Fault was also found with the opposing *eternalist* position. Temporal relations cannot be regarded as *eternal* in the sense of existing at all times, because then time would be required for time. Events, on the other hand, exist neither timelessly nor at all times, but at the times of their occurrence.
- We considered Aristotle's solution to Zeno's Arrow Paradox, according to which the arrow can correctly be said to have moved through an interval even if it cannot be said to be moving in any instant of its motion; and found that the same conclusion applies on Russell's at-at theory of motion: the reality of passage stands or falls with the reality of motion.
- Finally we examined the arguments against passage based on the denseness property of the continuum, according to which no two point-events can be next to one another, so there can be no passage to one point-event from an immediately preceding one. It was argued that this no more precludes passage from an initial to a final state than it does motion from one point to another.

References

Aristotle. 1984. In *The Complete Works of Aristotle*, 2 vols., Revised Oxford Translation, ed. Jonathan Barnes. Princeton: Princeton University Press.

Arthur, Richard T.W. 1995. Newton's Fluxions and Equably Flowing Time. *Studies in History and Philosophy of Science* 28 (2): 323–351.

Arthur, Richard T.W. 2012. Time Atomism and Ash'arite Origins for Cartesian Occasionalism. In *Asia, Europe and the Emergence of Modern Science*, ed. Arun Bala, 73–92. New York: Palgrave Macmillan.

Augustine, St. 1993. *Confessions*. Trans. Frank Sheed. Indianapolis: Hackett Publ. Co.

Barbour, Julian B. 1999. *The End of Time: The Next Revolution in Physics* (2001 reprinting). Oxford: Oxford University Press.

Barbour, Julian B. 2001. *The Discovery of Dynamics*. An Edited Reissue of *Absolute or Relative Motion? Vol. 1: The Discovery of Dynamics*. Cambridge: Cambridge University Press, 1989. Oxford: Oxford University Press.

Barnes, Jonathan. 1987. *Early Greek Philosophy*. London: Penguin Classics.

Barrow, Isaac. 1670. *Lectiones Geometricae*. London.

Barrow, Isaac. 1916. In *Geometrical Lectures*. Translation of Barrow's *Lectiones Geometricae* [1670], ed. J.M. Child. Chicago and London: Open Court.

Čapek, Milič. 1971. *Bergson and Modern Physics*. Dordrecht: D. Reidel.

Cohen, Hermann. 1883. *Das Prinzip der Infinitesimalmethode und seine Geschichte: ein Kapitel zur Grundlegung der Erkenntnisskritik* [*The Principle of the Infinitesimal Method and its History: A Chapter toward the Foundations of the Critique of Knowledge*]. Berlin: Dümmler.

Cornford, Francis M. 1959. *Plato's Timaeus*. Indianapolis: Bobbs-Merrill.

Davies, Paul C. W. 1995. *About Time*. New York: Touchstone.

Descartes, René. 1985. In *The Philosophical Writings of Descartes*, vol. 1, ed. and trans. John Cottingham, Robert Stoothof, and Dugald Murdoch. Cambridge: Cambridge University Press.

Dieks, Dennis (ed.). 2006. *The Ontology of Spacetime*. Amsterdam: Elsevier.

Dorato, Mauro. 2006. The Irrelevance of the Presentist/Eternalist Debate for the Ontology of Minkowski Spacetime. In ed. Dieks, 93–109.

Duhem, Pierre M.M. 1985. *Medieval Cosmology: Theories of Infinity, Place, Time, Void, and the Plurality of Worlds*. Chicago: University of Chicago Press.

Eddington, Arthur S. 1929. *The Nature of the Physical World*. New York: Macmillan.

Friedman, Michael. 2000. *A Parting of the Ways: Carnap, Cassirer and Heidegger*. Peru, Illinois: Open Court.

Grünbaum, Adolf. 1950. Relativity and the Atomicity of Becoming. *Review of Metaphysics* 4 (2): 143–186.

Grünbaum, Adolf. 1971. The Meaning of Time. In *Basic Issues in the Philosophy of Time*, ed. Eugene Freeman and Wilfrid Sellars, 195–228. La Salle: Open Court.

Hinchliff, Mark. 2000. A Defense of Presentism in a Relativistic Setting. In *Philosophy of Science, vol. 67, Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers, Sept 2000*, S575–S586.

Jankowiak, Tim. 2013. Kant's Argument for the Principle of Intensive Magnitudes. *Kantian Review* 18 (3): 387–412.

Kneale, Martha. 1968–69. Eternity and Sempiternity. In *Proceedings of the Aristotelian Society*, vol. 69, 223–238.

Kneale, William. 1960–61. Time and Eternity in Theology. In *Proceedings of the Aristotelian Society*, vol. 61, 87–108.

Leibniz, Gottfried Wilhelm. 2001. In *The Labyrinth of the Continuum: Writings on the Continuum Problem.* ed., sel. & trans. With an Introductory Essay. R.T.W. Arthur. New Haven: Yale University Press.

Mach, Ernst. 1919. *The Science of Mechanics*. Translated by Thomas J. McCormack of Mach's *Die Mechanik in ihrer Entwickelung, historisch-kritisch dargestellt*, 1888. 4th edn. Chicago/London: Open Court.

Markosian, Ned. 2003. A Defense of Presentism. In *Oxford Studies in Metaphysics*, vol. 1, ed. Dean Zimmerman. Oxford: Oxford University Press.

Monier-Williams, James. 2007. Hermann Cohen's *Das Prinzip der Infinitesimal-Methode und seine Geschichte*. MA thesis, McMaster University.

Muir, John. 1861. *Original Sanskrit Texts on the Origin and Progress of the Religion and Institutions of India, Part III: The Vedas*. London/Edinburgh: Williams and Norgate.

Newton, Isaac. 1999. In *The Principia: Mathematical Principles of Mathematical Philosophy*, ed. and transl. I. Bernard Cohen, and Anne Whitman. Berkeley: University of California Press.

Ockham, William 1984. In *Quæstiones in Libros Physicorum Aristotelis, Opera Philosophica et Theologica*, vol. 6, ed. Stephan Brown. St. Bonaventure: Franciscan Institute.

Priest, Graham. 2006. In *Contradiction*, 2nd ed. First edition, Dordrecht: Nijhoff 1987. Oxford: Oxford University Press.

Rovelli, Carlo. 1998. Loop Quantum Gravity. *Living Reviews in Relativity* (Rrefereed Electronic Journal), Invited Review Paper. http://www.livingreviews.org/Articles/Volume1/1998-1rovelli; gr-qc/9709008.

Rovelli, Carlo. 2007. *Anaximander*. Yardley, PA: Westholme Publishing.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Russell, Bertrand. 1903. *Principles of Mathematics*. New York: W. W. Norton & Co.

Russell, Bertrand. 1912. The Philosophy of Bergson. *The Monist* 22: 321–347.

Russell, Bertrand. 1915. On the Experience of Time. *The Monist* 25: 212–233.

Russell, Bertrand. 1929. *Mysticism and Logic*. New York: W. W. Norton & Co.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Savitt, Steven F. 2006. Presentism and Eternalism in Perspective. In ed. Dieks, 111–127.

Shimony, Abner. 1998. Implications of Transience for Spacetime Structure. In *The Geometric Universe Science, Geometry, and the Work of Roger Penrose*, ed. S.A. Huggett, Lionel J. Mason, K.P. Tod, Sheung Tsou, and N.M.J. Woodhouse. OUP.

Sider, Ted. 1999. Presentism and Ontological Commitment. *Journal of Philosophy* 96: 325–347.

Smart, J.J.C. 1949. The River of Time. *Mind* 58 (232): 438–494.

Smart, J.J.C. 1968. *Between Science and Philosophy*. New York: Random House.

Sorabji, Richard. 1983. *Time, Creation, and the Continuum: Theories in Antiquity and the Early Middle Ages*. Ithaca, New York: Cornell University Press.

Stein, Nathanael. 2016. Aristotle on Parts of Time and Being in Time. *Review of Metaphysics* 69 (3): 495–518.

Unger, Roberto Mangabeira, and Lee Smolin. 2015. *The Singular Universe and the Reality of Time*. Cambridge: Cambridge University Press.

Whitehead, Alfred North. 1925. *Science and the Modern World*. New York: Macmillan.

Whitehead, Alfred North. 1930. *Process and Reality*. New York: Macmillan.

Whitrow, G.J. 1961. *The Natural Philosophy of Time*. CA, London: Harper, Nelson.

Chapter 3
Modern Objections to Time's Passage

> *World history consists of actual concrete happenings in a temporal sequence; it is not necessary or possible that happening should happen to them all over again.*
>
> — D. C. Williams (1951, 464).

3.1 Introduction

The "now", it is said, is absent from the laws of physics. What should be inferred from this? Philosophers and scientists have taken up entrenched positions. On one side, we have those who say, so much the worse for the "now"—it is obviously purely subjective; on the other, those who say, so much the worse for science, which cannot represent so obvious a fact of phenomenal experience.

A peremptory resolution of this dilemma might go as follows: if Aristotle is correct, then the different "nows" are just the times at which particular events occur. We no more expect references to the times of particular events to feature in the *laws* of physics than we expect references to their locations—although of course we will need information about such concrete particulars in order to apply the laws of physics to those situations.

But to many philosophers that answer has seemed inadequate to our experience of time. There is a long tradition in philosophy of regarding the now as essentially subjective: it is in the now that we experience everything that goes on in the world. What we know is gained through experience, but experience itself is rooted in the present. Such a stance has been taken by philosophers of very different persuasions, as we shall see. Here we find not just phenomenologists alongside analytic philosophers in the tradition of Humean empiricism, but also philosophers such as Henri Bergson (1859–1941) and Alfred North Whitehead (1861–1947), who began as mathematicians but came to see the mathematical representation of time as inadequate. Bergson criticized the physicists' idea of time as constituting a "spatialization of time", to which he opposed his idea of *durée réelle*, a "lived time" that he equated not only with becoming, but also with life and consciousness, while Whitehead (as we have

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_3

seen) proposed that the units of becoming are "actual occasions", which physics can treat only after they have been constituted by experience.[1]

In this chapter I will begin by examining Bergson's claim that the mathematical representation of time is a flawed spatialization of time. The kind of thing he has in mind is well illustrated by H. G. Wells's presentation of time as "the fourth dimension" in his inspirational 1895 novel *The Time Machine*.[2] The idea of time as a fourth dimension is not in itself objectionable; in fact, it has appeared especially prescient in the light of Minkowski's spacetime formulation of Einstein's Relativity Theory thirteen years later. But Wells (1866–1946) construes time as a kind of space to be surveyed by a consciousness to which its contents are presented. Although such a conception of time is clearly philosophically untenable, I argue that Bergson's analysis of its shortcomings does not quite hit its mark. While it is true that the representation of time by a line has invited the misconception that temporal passage is unreal, the fault lies not in succession being represented as a linear order, but in inferring from this static representation of change that no change is thereby represented. For without passage there are no events to be ordered successively, and successions of events are what constitute change.

The representation of time as an ordering of events is what prompted Russell's account of the nature of change, the "static" theory which was discussed and rejected in Chap. 2. According to that account change is nothing more than *temporal difference*: a change occurs if the properties ascribed to an individual are different at one time than another. The inadequacy of this account of change featured prominently in McTaggart's famous argument for the unreality of time, to an examination of which I turn next. He maintained that the Russellian account—which following his lead came to be called the "B theory" of time—could not account for the very existence of change. That, he held, is only explicable by recourse to the "A theory", according to which becoming is comprised by events' changing what he called their "A-determinations", namely their changing from being future, to being present, to being past—a theory that he then found self-contradictory. But McTaggart's analysis is itself deficient, as I proceed to argue. An examination of its shortcomings leads to a rejection of two of its main premises: (i) that events can be regarded as positioned in time independently of their having occurred, and (ii) that becoming must be characterized in terms of events' changing their relations to the now.

The falsity of these premises, I contend, undermines not only the "A theory", but just as surely the "B theory" (at least, as it is presented by some of its main proponents). In the following sections I give a fuller discussion of those criticisms. In Sect. 3.4 I examine the "block universe" picture advocated by B-theorists, and distinguish the various theses that have been advanced under that rubric. I argue that, despite the failings of the A theory, the philosophical objections to the reality of becoming advanced by B-theorists are not sound; nor either is the "growing block"

---

[1] For an succinct treatment of Whitehead's philosophy of time, see (Hurley 1986).

[2] There are two very good editions of this work, one with excellent critical apparatus and forerunners of the novella edited by Harry M. Geduld (Wells 1987), and the other with important related texts, such as those by Hinton (1884) and Newcomb (1894), edited by Nicholas Ruddick (Wells 2001).

model of the universe championed by some as a compromise solution. In connection with the second shortcoming of McTaggart's argument—his analysis of becoming in terms of a relation to the now—both proponents and opponents of passage have, as mentioned above, construed the now in terms of an observer's experience. In Sect. 3.5 I give a fuller discussion of that view, including a critique of Grünbaum's notion of the "mind-dependence of becoming". I find this notion as untenable as the rival views of becoming put forward by James, Bergson, and Whitehead that it was intended to supplant. The idea that becoming implies or involves consciousness is, in my view, a mistake. The latter authors' analyses, however, do have the great merit of drawing attention to the extendedness of all actual events, and the abstract nature of point events, a consideration that is sorely underappreciated in the literature.

3.2 Wellsian Time Travel and the Spatialization of Time

The brunt of Bergson's criticism falls on a particular conception of time in physics that gained traction in the late nineteenth century, when time came to be regarded as "the fourth dimension", orthogonal to the three dimensions of space. The idea of a fourth dimension was much discussed in the last two decades of that century following the mathematician C. Howard Hinton's publication of his monograph *What is the Fourth Dimension?* in 1887. In explaining this notion, Hinton asked his readers to picture a being confined to a plane, for whom a square would appear as a solid that could only be entered by breaking through one of the four lines making up its boundary. Now, "if such a square were to pass transverse to his plane, it would immediately disappear" (Hinton 1887, 207). But a cube passing transversely through the plane at right angles to it will appear as a square that lasts for some duration, then disappears. Ramping up one more dimension, we may envision the motion of a cube of four spatial dimensions:

> Similarly, a four-dimensional cube, or, as we may call it, a tesseract, which is generated from a cube by a movement of every part of the cube in a fourth direction at right angles to each of the three visible directions in the cube, if it moved transverse to our space, would appear as a lasting cube. (207)

The trend of Hinton's speculations is to conceive of the human soul as "a four-dimensional organism" (24), a "being, capable in itself of four dimensional movements, but in its experiences through the senses limited to three dimensions" (20).[3] It was H. G. Wells who first forcefully identified the fourth dimension with time (as

---

[3] In reporting this I do not mean to dismiss Hinton as some crank. Among the many merits of his book is his insightful discussion of rotations in higher dimensions. After a certain rotation in 4-space, he writes, "The dimension which appeared as duration before will become extension in one of our known dimensions, and a dimension which coincided with one of our space dimensions will appear as duration." (Hinton 1887, 241). This is a remarkable anticipation of Hawking's hypothesis of imaginary time, if we make due allowance for the fact that, in Minkowski spacetime, the signature of the four-dimensional metric is such that what is on a par with a dimension of space is time multiplied by the imaginary number i. Hawking hypothesized that the problem of the origin of

opposed to spirit, life or other candidates mooted by his contemporaries) in his "scientific romance" *The Time Machine*. In the philosophical preamble to the story, Wells acknowledges his debt to Hinton, while presenting *time* as the obvious candidate for the fourth dimension:

> There are really four dimensions, three of which we call the three planes of Space, and a fourth, Time. There is, however, a tendency to draw an unreal difference between the former three and the latter, because it happens that our consciousness moves intermittently in one direction along the latter from the beginning to the end of our lives. … some people who talk about the Fourth Dimension do not know what they mean by it. It is only another way of looking at time. *There is no difference between Time and any of the three dimensions of Space except that our consciousness moves along it*. (Wells 2001, 60)

This is not quite the same as Hinton's view. There, what exists is a world of four-dimensional organisms, each moving along a fourth spatial dimension orthogonal to the three we experience. As it moves through our 3-space, it appears, changes, and disappears. But these changes are mere successive appearances of these permanent beings, as they are revealed to an experiencing subject:

> Passing to four dimensions and our space, we can conceive that all things and movements in our world are the reading off of a permanent reality by a space of consciousness. Each atom at every moment is not what it was, but a new part of that endless line which is itself. And all this system successively revealed in the time which is but the succession of consciousness, separate as it is in parts, in its entirety is one vast unity. (Hinton 1887, 26)

Wells's simplification is to single out one of Hinton's four spatial dimensions as time, orthogonal to the other three.[4] Now, instead of a subject moving in four dimensions of space, it moves in three dimensions of space and one of time. But because the time dimension is really a spatial dimension of a different kind, the subject may move forward or back along this dimension at different rates. In order to make sense of this idea of a motion in time at different rates, Wells appeals to Hinton's idea that time is the succession of appearances of the four-dimensional world to a subjective consciousness. When Wells's "Medical Man" objects that "you cannot move at all in Time, you cannot get away from the present moment", the Time Traveller replies: "that is just where you are wrong. … We are always getting away from the present moment. Our mental existences, which are immaterial and have no dimensions, are passing along the Time Dimension with a uniform velocity from the cradle to the grave" (Wells 2001, 62). In keeping with this, when his Time Traveller sets off into the future at a rate of around a minute per second, he sees the housekeeper Mrs. Watchett, who in her time took about a minute to move across the room to the garden door, "shoot across the room like a rocket" (76). When he increases his pace to "over

---

the universe in time could be dissolved by allowing that the (imaginary) time dimension rotates to become a dimension of space as we extrapolate back to the beginning of the universe. See Hawking (1988, 136–141). This idea spatializes time in precisely the sense that Bergson criticized.

[4] Actually, in connection with the fourth dimension Wells specifically mentions Simon Newcomb (Wells 2001, 61), rather than Hinton. But as discussed by Nicholas Ruddick in his introduction to his edition of *The Time Machine* (Wells 2001, 22–24), Wells was indebted to Hinton's essay for some of his ideas, as were both of them to Edwin A. Abbott's *Flatland* (1884).

a year a minute", he sees "minute by minute the white snow flash across the world and vanish, … followed by the bright, brief green of spring" (76).

This conception of time in the four dimensions of spacetime as being traversed by a moving consciousness has been surprisingly influential. Thus A. d'Abro wrote in 1950:

> When we consider the four-dimensional space-time continuum, where space and time are on the same footing, there is nothing to suggest either a flowing of time or a privileged direction for this flow. In order to conform the theory to the facts of experience, it is therefore necessary to postulate that our consciousness rises along the world-line of our body through space-time, discovering the events on its course. (d'Abro 1950, 206)

This same view had already been voiced by the mathematical physicist and philosopher of science Hermann Weyl: "Only to the gaze of my consciousness, crawling upward along the life line of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time" (Weyl 1949, 116).[5]

I say "surprisingly influential" because there are many fairly obvious objections to such a conception. In the first place, in Wells's story there is an evident discrepancy between his defence of time travel in terms of being conscious of events at a different time, and actually physically visiting that time with mind and body. He writes: "For instance, if I am recalling an incident very vividly I go back to the instant of its occurrence: I become absent-minded, as you say, and jump back for a moment. Of course, we have no means of staying back for any length of time…" (Wells 2001, 62). But recalling an incident in one's past is only bringing the memory of that event to present consciousness, which is obviously different from being physically transported there, body and soul, by a Time Machine.[6] There is also a discrepancy between the times in which the Traveller would be experiencing events normally, and the times in which he is experiencing the same events while travelling in his machine. If these times are the same (as they would be if times are identified by their experiential content) then there is no time travel; if they are different, then he is visiting times in a different time.[7] And if we take the latter seriously, then there are all those problems having to do with whether or not one can change the past—problems that Wells adroitly sidesteps by concentrating on his Traveller's visits to the future![8]

---

[5] Weyl had said almost exactly the same thing in his earlier *Mind and Nature* (Weyl 1934, 76), and earlier still in his "Time Relations in the Cosmos, Proper Time, Lived Time, and Metaphysical Time" of 1927; see (Weyl 2009, 135, 32).

[6] This criticism evokes, curiously, a similar criticism that Russell made of Bergson: "The whole of Bergson's theory of duration and time rests throughout on the elementary confusion between the present occurrence of a recollection and the past occurrence which is recollected." (Russell 1912, 342).

[7] The same point about Wellsian time travel was made eloquently by D. C. Williams: "Time travel, then, is analyzable either as the banality that at each different moment we occupy a different moment from the one we occupied before, or the contradiction that at each different moment we occupy a different moment from the one which we are then occupying—that five minutes from now, for example, I may be a hundred years from now" (Williams 1951, 463).

[8] As David Wright has reminded me (private communication), Wells's Traveller does return to the relative past from the far future of the blood-red dying sun, and would have fallen headlong into

But the most profound difficulty is in the notion of moving along the time dimension at all. Such a journey, as Larry Dwyer notes, would have to be represented on a space-time diagram not as a static line, but instead by a moving point (Dwyer 1975, 342). Yet "There is clearly no room in the space-time picture for movement through space-time", as Jack Smart observed. "What would movement through time be? Change of time with respect to what?" (Smart 1968, 256). The very idea of a consciousness or anything else "moving" along a time line must be rejected as pure confusion. Not that this precludes the possibility of any kind of time travel: in fact, as we shall see below, relativity theory makes one kind of travel into a future (where everything else has aged more than the traveller) an inevitable consequence of travelling back to one's starting point at high speeds; and in General Relativity the possibility is raised of travelling back in time in a spacetime that curves back onto itself. But, as Dwyer argues, if time travel is possible, then what occurs at any given point in spacetime can only occur once: it cannot be "visited again". And the root reason for this is that time is not, after all, a kind of space that can be visited by some detached consciousness.

It was perhaps as a result of his reflections on the speculations of such as Hinton and Wells that Henri Bergson came to a realization that was seminal for his philosophy of time. This was precisely that in physics time is treated as though it were a kind of space. The time co-ordinate is on a par with the three spatial coordinates, and the system is regarded as completely specified once the values of the variables describing it are specified at each spatial point and instant. Systems in the world are represented as configurations, whose different states are calculable from the initial conditions, given the appropriate laws. But such systems and their various states, regarded as laid out through time, are not accessible to immediate experience. Moreover, they do not change, any more than do the laws themselves.

This scenario is well summarized by Lee Smolin, who describes the implications of this way of viewing the world in terms of configurations, initial conditions and laws of motion expressed in equations, as "terrifying" (Smolin 2013, 44). For the laws themselves, "do not evolve, they simply are" (Smolin 2013, 51).[9] To the extent that the assumptions underlying this picture are physically realized, he argues,

> time is inessential and can be removed from the description of the world. If the space of possible configurations can be specified timelessly, and the laws can as well, then the history of any system need not be seen as evolving in time. … It is sufficient, for answering any question physics can pose, to see the whole history of any system as a single frozen curve in configuration space. The seemingly most essential aspect of our experience of the world—its presentation to us as a succession of present moments—is missing from our most successful paradigm for the description of nature. (Smolin 2013, 44, 51–52)

Similarly, when you represent motions by a graph of space against time, the result is a static drawing of lines in space. Once time is represented by a horizontal line, you

---

this paradox of changing the past if he had made good on his intention on the way back to save Weena from the Morlocks.

[9] Smolin is not advocating this way of viewing physics, just explaining it. In talking of configurations statically laid out, he probably has in mind the theory of Julian Barbour described in Chap. 2, and to which we return in Chap. 8.

can cast your gaze from left to right or right to left with impunity, just as you can with any line. The actual passing of time is not represented. So it is with spacetime. Once all the events in spacetime are represented on a spacetime diagram, and the structures of that spacetime are represented with timeless mathematics, where is it that the passing of time enters into physics?

Exactly these kinds of considerations horrified Henri Bergson in the late nineteenth century. To him they seemed to show a fundamental inadequacy in the mathematical representation of time in physics.[10] In his doctoral thesis, published in 1889 as *Time and Free Will*, Bergson (1910) argued that such a representation foists spatial concepts onto time, thus distorting its true nature. This true nature, Bergson claimed, is *duration*, by which he meant time as it is actually experienced, or "lived time". Duration is what constitutes reality, he explains in his *Creative Evolution*, and reality consists in a perpetual becoming (Bergson 1944, 297). So he is not using the term *duration* in the classical sense I outlined in the previous chapter; a better word might have been *transience*. He is urging us to recognize that we experience the world as a constant process of change, becoming, and that this just is what reality is. Everything else is abstraction and misrepresentation. "Things and states," he writes there, "are only views, taken by our mind, of becoming. There are no things, there are only actions" (270–1). This true duration contrasts with time as it is represented mathematically, where the intellect presents it as a succession of "freeze frames", as in a movie. This Bergson castigates as the "cinematographic view of change". "Every attempt to reconstitute change out of states," he writes, "implies the absurd proposition, that movement is made of immobilities" (334–35).[11]

Bergson's philosophy was subjected to searing criticism by Bertrand Russell in 1912. As we might expect given his account of motion,[12] he took exception to Bergson's presentation of mathematical time as cinematographic, and also to Bergson's appeal to Zeno's paradox of the moving arrow in support. "Mathematics conceives change, even continuous change, as constituted by a series of states; Bergson, on the contrary, contends that no series of states can represent what is continuous, and that in change a thing is never in any state at all" (Russell 1912, 338). This would be true, Russell admonishes Bergson, if mathematics represented motion as a discontinuous series of states. But it does not. "A cinematograph in which there are an infinite number of films, and in which there is never a *next* film because an infinite number come

---

[10] "It was the analysis of the notion of time as that enters into mechanics and physics," he is quoted as saying, "which overturned all my ideas. I saw, to my great astonishment, that scientific time does not *endure*. This led me to change my point of view completely" (*Encyc. Brit.* article on Bergson).

[11] Cf. Bergson's criticism of associationism, which "requires that each psychical state should be a kind of atom, a simple element." (Bergson 1913, 172): "The capital error of associationism is that it substitutes for this continuity of becoming, which is the living reality, a discontinuous multiplicity of elements, inert and juxtaposed." (171).

[12] "Motion is the occupation, by one entity, of a continuous series of places at a continuous series of times" (Russell 1903, §442, 469).

between any two, will perfectly represent a continuous motion.[13] Wherein, then, lies the force of Zeno's argument?" (339) The source of Bergson's misconception that mathematics represents motion as compounded from immobilities, Russell contends (ironically), lies in his view that "mathematical time … is really a form of space" (327); that the intellect, in representing time, necessarily spatializes it. Bergson, Russell charges, mistakes an intellectual idea for an image, and consequently does not appreciate the element of abstraction essential to mathematics. If we represent an ordering of states in time, we have abstracted the notion of an asymmetric order from the events themselves. Temporal precedence will now share the same abstract character as an asymmetric ordering of points, say, from left to right. Provided the points have the same order as the instants, we will not have misrepresented this property of time by depicting them spatially. Analogously, "the twelve apostles, the twelve tribes of Israel, the twelve months, the twelve signs of the zodiac, are all collections of units, yet no one of them is the number 12, still less is it number in general" (335). The fact that instants are temporally ordered rather than spatially ordered makes no more difference to their order than the apostles being apostles rather than signs of the zodiac makes to their numbering 12.[14]

I think this last point is essentially correct. Time is not spatialized merely by the fact that we have abstracted certain features of it that it has in common with space, namely by representing it as involving an ordering of states. Bergson criticizes the mathematical representation of a process by a line as being static, and not changing. But when we represent a process as having taken place through a continuous sequence of states, and draw a line to represent its trajectory, we are representing it as *having become*. We have abstracted away from its actual becoming, and are representing it as a finished process, something that has become, has changed. Its having become or changed is presupposed in the representation, not negated by it. So the fallacy of the spatialization of time is not that time is necessarily misrepresented by being mathematized, as Bergson alleges. It is to infer from the static representation of change that no change is thereby represented.

Nevertheless, there is a sense in which the latter is precisely what Russell did when he claimed that Zeno was right to have inferred that change is illusory from the fact that his arrow is at rest at every instant of its flight—and this is certainly to construe movement as "made of immobilities", just as Bergson alleged. In fact Bergson makes an analogous criticism of Descartes for his claim that body consists in extension, completely independent of mind. In so doing, Bergson contends, Descartes attributes divisibility to bodies, when it is in fact an artefact of how he has represented them. He takes the qualities of the intellectual representation for qualities of what is represented

---

[13] To be pedantic, the infinity of frames between any two must be a non-denumerable one, on Cantor's account, otherwise this will only yield the property of denseness, and not full mathematical continuity.

[14] Despite his agreement with much of what Bergson says about time as rooted in lived experience, Whitehead agrees with Russell in this criticism: "in so far as Bergson ascribes the 'spatialization' of the worlds to a distortion introduced by the intellect, he is in error. This spatialization is a real factor in the physical constitution of every actual occasion belonging to the life history of an enduring physical object." (Whitehead 1930, 489).

in reality.[15] This is an instance of what Whitehead (under Bergson's influence) called the Fallacy of Misplaced Concreteness (Whitehead 1925, 64, 72). But the fallacy lies in taking the representation as the thing, and as being all there is to that thing in reality; not in abstracting certain features of the thing and modelling those mathematically. Descartes's mistake was in taking bodies to consist in pure extension because he could conceive that clearly and distinctly, not in the idea that bodies may be represented as extended, and thus made amenable (in this respect) to mathematization. Similarly, the fallacy of the spatialization of time lies not in representing it as a succession of states, but in inferring from the fact that the representation of succession is static, that no change is represented in it.[16] This fallacy is committed by all those who follow H. G. Wells in wanting to re-inject motion into a spacetime diagram, and have a point of consciousness move up along a worldline.

So we can certainly sympathize with Bergson if we remember that the kind of static view he is opposing is the very view that Russell had outlined in his *Principles of Mathematics*:

> every term is eternal, timeless, and immutable; the relations it may have to parts of time are equally immutable. It is merely the fact that different terms are related to different times that makes the difference between what exists at one time and what exists at another. (Russell 1903, 471)

This is a paragon of the *eternalist* philosophy we found fault with in Chap. 2. The fact that the Great Fire of London occurred before the rebuilding of St Paul's Cathedral, for instance, is indeed irrevocable or immutable, once the events have occurred. But the "terms" that are related as one before the other are *events* that occurred precisely when they did indeed occur (1666 for the Great Fire, 1675–1711 for the rebuilding). These events occurred in time, not timelessly.

Moreover, if we concentrate our attention for a moment on such real events as these, we get a clearer insight into Bergson's position. The Great Fire and the rebuilding of St. Paul's are both continuously unfolding events: their constituents are other extended events, such as stones being cut and set in place, or Wren being winched up to the top of the dome to inspect progress. When Bergson spoke of the continuity of time, it was such interpenetrating and imprecisely defined events as these that he had in mind, not the series of externally related point-events that Russell proposed. The spatialization of time, for Bergson, consists in taking that model of temporal conti-

---

[15] Cf. Bergson: "But suppose now that this homogeneous space is not logically anterior, but posterior, to material things and to the pure knowledge which we can have of them; … suppose that homogeneous space concerns our action and only our action, being like an infinitely fine network which we stretch beneath material continuity in order to render ourselves masters of it, to decompose it according to the plan of our activities and our needs." (1913, 307–8). I am much indebted to my colleague Barry Allen both for this reference and for correcting some misconceptions I had of Bergson's views.

[16] Steven Savitt has made the same objection. Anyone asking where change is represented on a spacetime diagram "is confusing a static representation with a representation of stasis." (Savitt 2002, 162–3). Cf. Lee Smolin: "the fallacy of the spatialization of time … is a consequence of forgetting the distinction between recording motion in time and time itself." (Smolin 2013, 35).

nuity—a model that reduces time to external relations among abstract point-events and excludes passage—and mistaking it for the real thing.[17]

This brings us to another aspect to Russell's static view. This is the idea that change consists in the mere difference in some property at different times. Jack Smart seems to subscribe to this view in his 1949 article, "The River of Time". Discussing events such as "the 3-dimensional shape of a man at any instant of his life", Smart writes:

> if an event (in this sense) may be said to be susceptible of change, say along the time-dimension, this is a usage very different to that in which we say "the traffic light changed". It is analogous to that in which we say "the country changes as you go north" and has nothing to do with our present puzzle at all. The country does not "really change", i.e., it does not change in the sense in which the traffic light does. It is just different in one place from what it is at another. (Smart 1949, 490)

But if this is all there is to change, difference between states of affairs at one time and another, with no passage from earlier to later, does anything really change? Or is the idea of the passage of time just a facet of our subjective experience? This uncomfortable dilemma formed the starting point for McTaggart's celebrated argument for the ideality of time, published in 1908. He claimed that neither theory can be sustained—neither Russell's static theory, nor the idea of explaining change in terms of a changing relation of events to the experienced now.

3.3 McTaggart's A- and B-Series

John McTaggart Ellis McTaggart (1866–1925) was an idealist, and he offered his famous argument for time's unreality to buttress that position. "I believe that nothing that exists is temporal," he wrote in his *The Nature of Existence*, "and that therefore time is unreal" (McTaggart 1927, §304, 9). He began by imagining that all events, past, present and future, are given in their temporal relations: for instance, the partitioning of India precedes the assassination of Gandhi, the publication of the Communist Manifesto occurs after the births of Marx and Engels, and so forth. More precisely, he began by assuming that the "series of positions in time which runs from earlier to later" is given, a series he dubbed "the B series", and that "the contents of a position in time are called events" (McTaggart 1908, 458). He also assumed another series, which he called "the A series", the "series of positions running from the past to the present, and then from the present to the near future and the far future". This series differs crucially from the B series in that it is only in this series that the "now" is represented, and without the now changing its position with respect to the events in the A series, the same event could not become present from having been future, or past from having been present.

Relative positions in the B series, McTaggart claimed, "are permanent, while those of the latter are not". He explained, "if M is ever earlier than N, it is always earlier than N. But an event, which is now present, was future and will be past." McTaggart

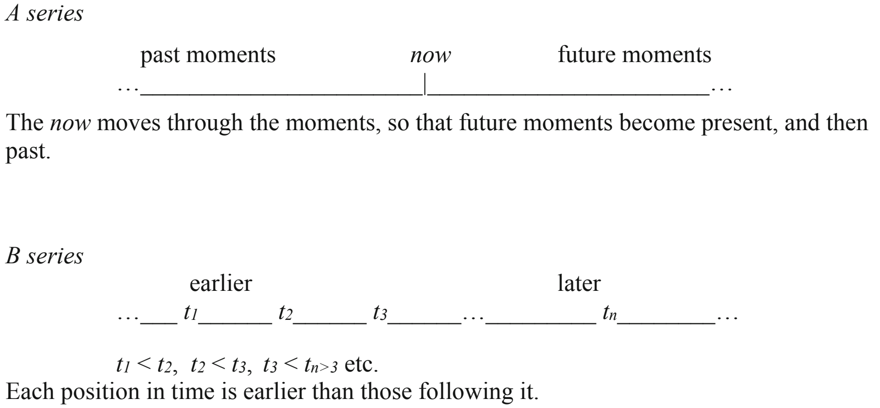
---

[17] Here again I am indebted to Barry Allen for his sage advice on Bergson.

calls each position in time a *moment*. Thus the relative positions of events in the B series, its moments, are permanent, but these moments change their (absolute) positions in the A series, from future to present to past. McTaggart grants that the distinctions of positions in the B series, being permanent, "might be held to be more objective, and to be more essential to the nature of time" (458). But he regards "the distinction of past, present and future" to be not only "as essential to time as the distinction of earlier and later", but, "in a certain sense … more fundamental" (458). This, he claims, is because the positions in the B series are permanent: there is no change there. That the partitioning of India precedes the assassination of Gandhi is not a fact that will change in time. But without change there cannot be any time. And change is only found, on this analysis, in the changing positions of the events in the A series, which, having been future, slip through the present into the past. In the more famous and much analysed part of his argument, McTaggart then proceeds to find a contradiction in the idea that the moments (positions) in the A series can in fact change.

Although reactions to McTaggart's argument have been many and varied, modern philosophers of science have tended to accept the basic structure of his argument. They have assimilated his critique of the A series to their own criticisms of the moving now. If events are changing their A-determinations, i.e. changing from future, to present to past, then their positions in the A series (their "moments", in McTaggart's idiosyncratic terminology) are changing. But any such changes will have to occur in time. Now moments cannot change in time (events cannot change their positions in time) without presupposing another time in which that change occurs (Fig. 3.1).

In fact, this latter line of argument has seemed so persuasive to some that they have grabbed the bull by the horns: why should there not be more than one dimension of time? Notorious in this respect is J. W. Dunne, who in 1927 wrote a book, *An*



The *now* moves through the moments, so that future moments become present, and then past.

$t_1 < t_2,\ t_2 < t_3,\ t_3 < t_{n>3}$ etc.

Each position in time is earlier than those following it.

Fig. 3.1 McTaggart's A and B series

Experiment with Time, in which he advocates just this.[18] He notes that "like Hinton, Wells fails to mention that anything that moves in Time must *take Time* over its movement" (Dunne 1938, 144); so he introduces an extra time dimension in which this movement can take place. Acknowledging that the same problem will recur for this second time dimension, he simply cedes the point: "*Every Time-travelling field of presentation is contained within a field one dimension larger, travelling in another dimension of Time, the larger field covering events which are 'past' and 'future', as well as 'present', to the smaller field*" (187). He is also explicit that each of these "fields of presentation" involves an observer, with the consequence that the series of fields of presentation must therefore involve "*the existence of a serial observer*" (188). Dunne's idea is thus that each field of presentation is an ordering of events from past to future, with the motion from past to future being surveyed from the next larger dimension of time; and since there must be an observer to whom each field is being presented, he hypothesizes a "serial observer" to whom all these fields are presented. Thus Dunne proposes a "serial time",[19] in which each field of presentation is travelling in a larger field where past, present and future are laid out extensively, as positions in a space that can be travelled through. In so doing he commits the same fallacy as had Wells in treating time as a space that can be traversed. Dunne criticizes Bergson for his idea that "pure duration", as "time flowing", will successively add to "time flown", observing that "His growing 'past' takes Time to grow" (155). But he fails to see that his own solution involves the same fallacy that Russell had already identified in Bergson's view, namely the erroneous supposition that the representation of the "extensive" aspect of time as an asymmetric ordering of events makes time into a space that can be moved over.

The untenability of Dunne's particular attempt does not refute the whole idea of time having two (or more) dimensions, and the notion of a two-dimensional time has continued to intrigue philosophers and physicists alike—especially given the embroilment of time with space made evident in relativity theory.[20] But insofar as these attempts depend on a distinction between an extensive time and a time in which some now or consciousness navigates that extension, they are all committing

---

[18] Although Dunne is explicit in acknowledging his sources, he does not mention the idealist philosopher Bradley. But according to McTaggart (he does not give references), "The second objection is based on the possibility, discussed by Mr. Bradley, that there might be several independent time series in reality. For Mr. Bradley, indeed, time is only appearance. There is no real time at all, and therefore there are not several real series of time. But the hypothesis here is that there should be within reality several real and independent time-series." (McTaggart 1908, 466).

[19] There's a potential for some confusion here. By "serial time" Dunne means his infinitely regressing series of time dimensions. This should not be confused with the "serial order" of events in time (referred to, for example, by Russell) . This is what is now usually called a "*strict total ordering*", denoting a relation that is asymmetric, transitive, and simply connexive; that is, for any three events x, y and z, if x is before y then y is not before x; if x is before y and y before z, then x is before z; and if x and y are different events, one is before the other. As we shall see in Chap. 6, this last property does not hold for point-events "being in the absolute past of" one another in special relativity, which is instead a *strict partial ordering*.

[20] Notably, however, even (bosonic) String Theory, with its posited 25 dimensions of space, still has only one of time.

the same fallacy as that which undermines Wells's notion of time travel and Dunne's "serial time". As Jack Smart has observed, the "now" can seem to us like "the crest of a wave on which we (like surfers) are being rushed toward into the future." But "no wave can flow upwards through spacetime, since time is already there in spacetime" (Smart 1968, 256).

So, many philosophers have argued, the whole idea of events coming to be by taking on differing determinations of future, present and past must be abandoned. To suppose that events can change their positions relative to the now is just as contradictory as to suppose that the now itself can move. As D. C. Williams has argued, "Nothing can 'move' in time alone any more than in space alone, and time itself can not 'move' any more than space itself" (Williams 1951, 463). When passage is so conceived, it "incomprehensibly doubles its world by re-introducing terms like 'moving' and 'becoming'", and leads to the "preposterous" idea of time's motion that is at the root of the infinite regress embraced by Dunne (Williams 1951, 463–4).

But what about the "now" on this view? Don't events change their status in relation to *now*? "The term 'the present'", Williams explains, "is the conventional way of designating the cross-section of events which are simultaneous with the uttering of the phrase…" (463). Smart concurs:

> When we say that an event is in the future we are saying that it is later than our utterance, when we say it is present we are saying that it is simultaneous with our utterance, we say that it is past we are saying that it is earlier than our utterance. (Smart 1968, 255)

Thus statements about what is past, present or future make an essential reference to the time at which they are expressed or uttered. This is called the *indexicality* or *token-reflexiveness* of the *now*: past, present and future are indexed to the time at which they are uttered; what is past, present or future refers back to the token of expression or speech act in which it occurs.[21] On this reading, past, present and future become, as it were, accidental features of time, rather than constituting it in an essential way. McTaggart's "A-determinations" are rendered irrelevant to the reality of time, which is held to be constituted by the temporal relations of the B series alone.

Thus, concerning the first part of McTaggart's argument, these philosophers deny that without passage or becoming there would be no change. "There is passage", Williams contends, "but it is nothing extra" (Williams 1951, 463).

> Time 'flows' only in the sense in which a line flows or a landscape 'recedes into the west.' That is, it is an ordered extension. And each of us proceeds through time only as a fence proceeds across a farm: that is, parts of our being, and the fence's, occupy successive instants and points, respectively. (463)

D. H. Mellor also subscribes to this view in his *Real Time* (Mellor 1988). He gives an analogy with a poker with one end in a hot fire (Mellor 1988, 84). Just as the

---

[21] The notion of indexicality was introduced by Charles Sanders Peirce (see the term 'Index' in the Digital Companion to C. S. Peirce at http://www.commens.org/dictionary/term/index); the notion of token-reflexiveness is due to Reichenbach (see the Oxford Reference under 'token-reflexive': http://www.oxfordreference.com/view/10.1093/oi/authority.20110803104818996).

temperature of a poker may be said to change along its length, so temporal change may be understood in terms of temporal difference: a velocity, for example, may be said to change over a given time interval if its value increases across the interval, in the same way that the temperature of a poker may increase toward the end that is in the fire. Thus these authors endorse McTaggart's fleeting suggestion that one might take the B series to be "more objective, and … more essential to the nature of time" than the A series—overriding his objections to the coherency of this view—while agreeing with him that the A series is self-contradictory, and temporal becoming a mere illusion. This is the so-called "B theory" of time, usually assimilated to the static view.

In all this, however, there is a failure to recognize just how precarious are the foundations on which McTaggart's structure is built. For his initial supposition that events are "given" in their (B series) temporal relations of "earlier" and "later" is not as innocent as it may seem. You can imagine all events and their temporal relations *sub specie aeternitatis*, as they used to say, "from the point of view of eternity". This is to imagine them as existing timelessly. McTaggart, however, characterizes them as "permanent", "always the same", unchanging. Now, if something is to be *permanent*, it must stay the same over time. It can only be true that event M "is *always* earlier than N" if there is a time in which the temporal relations between these events could conceivably change. But this is to assume a time outside time, as was argued in Chap. 2. Russell prefers to talk of "terms" of temporal relations. But as we have seen, for him, too, both the terms and the relations among them are classed as "immutable", "eternal" (Russell 1903, 471).

This criticism deserves a more detailed analysis, given its significance for the origins of the B theory. For it is not just the assumption of the timelessness of its terms that is a difficulty at the heart of the B theory, but also the theory of change on which it is based, namely Russell's static theory of change as temporal difference. To see this it is worth looking at McTaggart's criticisms of the B theory in his posthumously published work (McTaggart 1927). As McTaggart notes there, the thesis of the indexicality of the *now* has its origin with Russell, who began his important paper "On the Experience of Time" (Russell 1915) by making this very point. But given Russell's construal of change as temporal difference, McTaggart argues, this construal of the now as indexical is impotent to save the reality of change in the B series by itself. I think he is right about this, and it is worth pursuing for the light it throws on the B theory and its constituent assumptions.

Change, according to Russell's theory, consists in a difference "in respect of truth or falsehood" between a proposition about an entity at some time t and a second proposition about the same entity at another time t' (Russell 1903, 469). To use McTaggart's example, "my poker is hot at 8 p.m. on Monday, 31 October, 2016, but not hot on at 8 p.m. on the following day." But, McTaggart observes, although this makes for a change in the poker, there is no change in the qualities of the poker. "The fact that it is hot at one point in a series and cold at other points cannot give change, if neither of these facts change—and neither of them does" (McTaggart 1927, §315, 16). The problem is that Russell has characterized change in terms of a difference between the truth value of a proposition about an entity (here, the poker) at one time and its truth at another, in other words, in terms of facts about its properties at those times. But these facts do not change. "Nor," McTaggart adds, "does any other fact about the poker change, unless its presentness, pastness, or futurity change" (16).

That remark suggests that an appeal to the A-determinations might save the reality of change. But, given Russell's construal of A-determinations as indexical or token-reflexive, it does not. McTaggart gives as an example the Battle of Waterloo, which changed from having been an event in the future to being one in the past. Now, according to Russell's theory of A-determinations, this means merely that "the battle of Waterloo is earlier than this judgment" (16). But such statements "are either always true or always false", so that again, no facts actually change.

Now McTaggart agrees with Russell that the idea of A-determinations changing is fraught with contradiction. (I will not go into his arguments here; suffice to say that he recognizes that it leads to an infinite regress.) But as we have just seen, he insists that without appeal to them, there is no change at all. The B series, he argues, is unintelligible without "time-determinations", such as "is past" or "is yet to come".[22] "So it follows that there can be no B series where there is no A series, since where there is no A series there is no time."

Nevertheless, he maintains, if the A-series is discarded, the events are still given in order:

> But it does not follow that, if we subtract the determinations of the A series from time, we shall have no series left at all. There is a series—a series of the permanent relations to one another of those realities which in time are events—and it is the combination of this series with the A determinations which gives time. ... But this other series—let us call it the C series—is not temporal, for it involves no change, but only an order. (McTaggart 1927, 461)

Thus, having proved to his own satisfaction that the B series depends on the A series, and that the latter is fraught with contradiction, McTaggart arrives at a series of terms—"those realities which in time are events"—but which are related by an asymmetric relation that is not temporal. There is no time in the C series. Although he does not claim to have proved the reality of the C series, he thinks "there are good reasons for supposing that such a C series does actually exist, in every case in which there is the appearance of a time-series" (§347, 30). This he takes as a vindication of the Hegelian view: "For Hegel regarded the order of the time-series as a reflection, though a distorted reflection, of something in the real nature of the timeless reality..." (§350, 31).

Now one might object that McTaggart was mistaken in looking for change in a change of facts, and that Russell was correct in viewing propositions as timeless. But contrary to what both authors believed, this does not make them eternal in the sense of "either always true or always false", since that little word "always" presupposes another time in which temporal relations might change, which we have seen to be in

---

[22] "For in order to get change," McTaggart argues, "and change in a given direction, it is sufficient that one position in the C series should be Present, to the exclusion of all others, and that this characteristic of presentness should pass along the series in such a way that all positions on the one side of the Present have been present, and all positions on the other side of it will be present" (McTaggart 1927, 463).

error. It is simply a mistake to look for change in changing facts about events, as if the events (or terms) are already "given", and something must happen to them. This is just the wrong place to look for change and becoming. The events *are* the changes, but they do not exist until they occur.[23]

Nevertheless, McTaggart points up a real difficulty with Russell's theory of change as temporal difference. McTaggart analyses becoming in terms of a changing relation to the now, which he rejects as contradictory. So does Russell; but in retaining the idea that events are given in an asymmetrical order, while insisting that there is no transition from earlier to later, Russell has cut out the ground for associating the sense of the ordering of events with the direction in which they change. Given Russell's static analysis of change and his denial of the A-series, McTaggart is arguing, there is no ground for identifying the resulting ordering of events as a B series rather than a C series, a merely logical ordering of terms in an asymmetric series. "The C series will include as terms everything which appears to us as an event in time, and the C series will contain the realities in the same order as the events are ranged in by the relations of earlier and later" (§351, 31). An event *a* that is earlier than an event *b* in the B series will be earlier than it in the C series; but there is no more a transition from *a* to *b* (rather than *b* to *a*) in the B series than there is in the C series. Although the B series is a series of changes, "and these changes go in one direction and not in the other" (§698, 347), McTaggart objects, such a fact is not represented in Russell's static theory of change.[24]

The neo-Hegelian assumption that Russell unwittingly shares with McTaggart, is that because events are the "terms" of temporal relations, they exist.[25] This amounts to the claim that the real is what exists as an object of thought—since we can think of future events, they are real. The fact that Russell and McTaggart conceive of events as *already given in their succession* is what makes it impossible for them to coherently understand becoming.[26] Secondly, they are both committed to a view of temporal change as involving no passage from the earlier to the later. But the point is that if

---

[23] McTaggart's contention that events in his 'C' series are "permanent" is also objectionable, since again this would require a time in which they would remain the same, contrary to the alleged timelessness of the series. If these events are not ordered in temporal succession, moreover, it is impossible to see what their order is supposed to consist in.

[24] Here one might argue that, once passage is denied, there is equally no basis for the asymmetric ordering McTaggart supposes for his C series. If a given event *a* comes after another *b*, then *b* cannot come after *a*. This gives an asymmetric ordering. But if events are just 'at' their positions in time, as Russell and McTaggart assume, then we simply have a set of events with no natural ordering, not an asymmetric ordering with two different senses. I will return to this issue in Chap. 4.

[25] The Hegelians had rejected the reality of number, space, time and matter on the grounds that they were internal relations, and therefore contradictory. Rejecting such relationalism at the turn of the century, Russell came to believe in a direct realism, where all relata exist along with external relations among them. "I began to believe everything the Hegelians disbelieved," he reports. "This gave me a very full universe. I imagined all the numbers sitting in a row in a Platonic heaven. I thought that points of space and instants of time were actually existing entities …" (Russell 1959, 48).

[26] Although I attribute this to Russell's and McTaggart's baptism in Hegelianism, there are many thinkers who are by no means Hegelians who think it natural to regard events as existing independently of their becoming. Thus A. N. Prior (1968, 1–2) refers to the "becoming ever more past" of

there is no passage of time, no transition from the earlier events to the later ones, then there is no change *in a given direction*, as McTaggart objected to Russell. Indeed, there are no events, since the happening of events in succession is what constitutes passage. We can *consider* events as laid out in succession, we can model them in spacetime. But existence in our consideration or in our model is not the same as the existence which is constituted by their actually occurring when they do.

The bottom line is that the B theory perspective, the "static" view, is just as problematic as the A-theoretic view. The A theory assumes that events can change their temporal relations, the B theory insists that these relations remain the same. But both views illicitly presuppose a time during which the relations either change or remain permanent. Both perspectives begin by assuming that all events "already" exist, and ask what more is required. The A theorist says they must become "now" at some time; the B theorist says that if they already exist, there is no need for them also to become. But the fallacy lies in interpreting the existence involved in assuming that "there are events" as a temporal existence. It is not. There are not future or past events in the sense that they exist now, i.e. at the time of my saying this. *The existence of things in time is their existence at the times of their occurrence*.

3.4 The Block Universe

Seduced by the simplicity of 'A' versus 'B', people are wont to call anyone who asserts the reality of becoming an A theorist, and anyone who believes that temporal relations do not change a B theorist. But if we want to stay true to McTaggart's framework, it is more accurate to say that an A theorist is one who believes that the reality of becoming is constituted by changing relations of events to the now; and a B theorist is one who believes that time is constituted by the succession of events without any objective passage or becoming. I do not think this exhausts the alternatives, since I find both conceptions untenable, as I have just argued. But to make the situation clearer, we need to say more about the B theory in its contemporary manifestations.

Proponents of the B theory often call it the "theory of the manifold", or the "block universe" view—a term apparently originating with William James.[27] But such expressions can mean different things to different people. (1) What people usually intend by the "block universe" is what James had in mind in coining the expression: a manifold of events that are just given statically, with becoming relegated

---

events, and Norton (2010, 24) writes that future events "become present" and then "drift off into the past", as Oliver Pooley notes (2013, 322). Although Pooley defends the objective passage of time, he endorses this conception of passage, claiming that it is a key challenge for the B theory to explain "why we are inclined to take the 'becoming more past' of events as an objective feature of reality" (Pooley 2013, 326).

[27] Cf. Russell on Bergson: "His whole doctrine of time is necessary for his vindication of freedom, for his escape from what William James called a 'block universe', for his doctrine of a perpetual flux in which there is nothing that flows" (Russell 1912, 34).

to how events appear to a subjective observer. (2) Others, disturbed by the fact that such a picture would seem to rule out free will and the contingency of future events, have embraced the "growing block" conception, where the only events that exist are past and present ones, leaving an "empty future". (3) In a third (unobjectionable) way of conceiving it, the block is a complete four-dimensional entity, and as such includes all events that ever were or will be, and therefore all processes of becoming, viewed as objective successions of events. Let us examine these three conceptions in turn. In the process, we will have to examine the viability of the notion of "tenseless occurrence", which has been given in defence of version (1), and tends to blur its distinction from (3).

The first way of conceiving it corresponds closely with the account of the B theory I gave in the previous section. So we find the same philosophers embracing it. "The universe", writes Jack Smart, "is a four-dimensional space-time manifold. Present, past and future are all equally real" (Smart 1968, 255). Similarly, D. C. Williams writes in his classic paper, "I believe that the universe consists, without residue, of the spread of events in space-time… The theory of the manifold is the very paradigm of philosophic understanding" (Williams 1951, 458, 471). On this view, "all events—-past, present and future—are equally real" (Davies 1995, 260). That some events are occurring *now*– means only that those events are occurring contemporaneously with the utterance of that observation, in keeping with the indexicality of the now discussed above. But since it is equally true of any event that it is happening now at the time of its occurrence, this does nothing to mark out any one event from any other. Therefore, it is inferred, passage or becoming present is not a feature of objective reality. "Everything—past, present and future—is there at once" (Barbour 1999, 143). Events simply *are*, and do not need also to "become". Their argument, in a nutshell, is this: if we assume that the universe is a four-dimensional space-time manifold and that this manifold is real, then the reality or existence of an event simply consists in its being contained in this manifold. It is therefore quite unnecessary to suppose that it also "becomes" or "becomes present". If an event already exists, it does not also need to come into being.[28]

There are two features of this argument on which I wish to concentrate: first, the sense in which it can truly be said that all events in the manifold are equally real; and secondly, once this sense is clarified, whether the inference to the unreality of the becoming of events can be sustained. I shall argue that it cannot, that the inference depends on a certain equivocation on the sense in which events can be said to exist.

Here again we need to be very careful about the slippery word "exists". As noted in Chap. 2, there are many senses of the words 'exists' and 'is' that can be distinguished. For current purposes, the main ones to consider would appear to be these: (i) to exist

---

[28] An explicit version of this argument is given by Craig Callender, in the course of criticizing so called "hybrid theories": "Because [upholders of] hybrid theories accept that a four-manifold is the arena of world history, they cannot—on pain of incoherency—analyze becoming in terms of the coming into existence of events. It simply doesn't make sense to say an existent event comes into being" (Callender 2000, S590). As Steven Savitt observes, however, it is perfectly coherent for an event to exist in the spacetime manifold (as "eternalists" insist) and yet "to occur at its allotted instant" or *now*, as the presentists insist (Savitt 2006, 126).

atemporally, as in ‘7 is prime’;[29] (ii) to exist at a given time or spacetime location; (iii) to exist at all times, or sempiternally; and (iv) to exist for a certain duration.

Now consider a point-event *a*. What does it mean to say that this event exists or is real? A straightforward answer would be this: an objectively existing event is whatever occurs at the place and time at which it is represented to occur, independently of anyone’s subjective experience. This involves existence in sense (ii); for point-events, clearly, senses (iii) and (iv) do not apply. At any rate, concerning the claim that all events in a spacetime manifold exist or are equally real, we can say that this is so in sense (ii): each of them is represented as being real, in the sense of occurring at the particular location in spacetime it occupies, independently of anyone’s experiencing it.

This will not, however, license an inference to the claim that all the events are *already* real. For such a claim makes an implicit reference to the time at which the event is being represented. (By the word ‘represented’ I mean ‘considered’, ‘spoken of’, ‘pictured in a spacetime diagram’, etc. I am not using it in any obscure technical sense). That a future event is represented as existing obviously does not make it exist at the time it is being represented. This point is granted by both Smart and Williams. Says Williams of his “theory of the manifold”, it “does not assert, therefore, that future things ‘already’ exist, or exist ‘forever’.” (Williams 1951, 470); says Smart, “Of course it could be misleading to say that according to the theory of relativity the future is ‘already in existence’” (Smart 1968, 226). Yet if there is no sense in which a given event “already” exists, it is hard to see the argument for the non-necessity of an event’s becoming, which earlier I summed up on their behalf in the words: “If an event already exists, it does not also need to come into being.”

Nevertheless, according to Smart and Williams there is an appropriate sense in which an event exists, namely in its being contained in the four-dimensional manifold. It IS in the manifold, where I have put the ‘IS’ in small capitals to denote that we are now using the atemporal ‘is’, the ‘is’ of sense (i) above. Thus if the manifold can be said to pre-exist in some sense, this will license an inference to the pre-existence of any of the events in it. But the conclusions we reached about the temporal existence of a singular event must apply a fortiori to the four-dimensional manifold. If future events do not exist at the time they are being represented, then the whole spacetime manifold cannot be said to exist then either. The spacetime manifold cannot be thought of as a thing existing on a par with three-dimensional physical objects, which exist through time. To suppose that a four-dimensional object has this sort of existence is to commit a paralogism.

There is a valid sense, however, in which we want to say that the spacetime manifold exists over and above the events in it. To say that a spacetime manifold exists objectively is to say that the metrical, topological and ordering relations among the events ARE as depicted, where the word ‘ARE’ is here being used atemporally, in

---

[29] I believe it can be seriously questioned whether the ‘is’ in ‘7 is prime’ is an ‘is’ of existence. It appears rather to be an ‘is’ of predication, which does not occur in a language like Kiswahili. But I am allowing it here on the principle of charity. It is usually referred to as connoting “tenseless existence”, on which more below. Savitt (2006) also explores other possible meanings of ‘is’, including the “detensed” ‘is’, where “x Is Φ” means “x either was, is or will be Φ”.

sense (i) above.[30] It is the copula we use to assert facts, and is not to be confused with the 'are' used to express duration in time. In the same way, if we say that event *a* IS before event *b*, we are stating a fact about their temporal relation. But it is a fallacy to speak of this relation as never changing, as being "immutable" or "permanent", as we saw Russell and McTaggart doing,[31] since these things can only be said of things existing in time. Neither point-events, nor temporal relations connecting them, nor four-dimensional objects like worldlines or indeed the whole of spacetime, can be said to *exist through time* (for a duration, or forever—senses (iii) and (iv) above), and only some events (a proper subset of those in the manifold) *exist at any given time* (sense (ii) above). One can grant that events EXIST in the sense of being contained in a manifold; but since a manifold can also only be said to EXIST in an atemporal sense (sense (i) above), we have not succeeded in identifying any sense of 'exist' that will support the argument that since events already exist, they do not need also to become.

At this point we should take note of a distinction that advocates of this first (static) conception of the block make in its defence. This is the distinction, proposed by Grünbaum and supported by Smart, between a becoming present of events and their "tenseless occurrence". That is, events are supposed to "occur" without needing to "come into existence".[32] Becoming present is identified as "happening in the tensed sense", in contrast to "occurring in the tenseless sense", which is equated with EXISTING in the atemporal sense discussed above. Happening in the tensed sense, according to this proposal, requires a conscious mind: becoming present is therefore not something that pertains to events in themselves. "Becoming is mind-dependent," Grünbaum writes, "because it is not an attribute of physical events per se, but requires the occurrence of certain *conceptualized conscious experiences* of the occurrence of physical events" (Grünbaum 1971, 197). While this mind-dependence thesis "does deny that physical events themselves happen in the tensed sense of coming into being apart from anyone's awareness of them", Grünbaum explains, it nevertheless "clearly avows that physical events do happen independently of any mind in the tenseless sense of merely occurring at later clock-times in the context of objective relations of earlier and later" (1971, 213–214).

We will come back to the mind-dependence of becoming below. For now I wish to concentrate on the notion that events are susceptible to two sorts of occurrence, "tensed" and "tenseless"—a distinction that has been widely adopted in the literature on time. I submit that this idea of "tenseless occurrence" is misconceived. It is one thing to talk of *verbs* being used tenselessly, as when Grünbaum claims that "to assert tenselessly that an event exists (occurs) is to claim that there is a time or

---

[30] This formulation, it seems to me, is fully in keeping with what Nerlich wants to say about the reality of spacetime and spacetime structure. See Nerlich (1994, 40ff).

[31] "If N is ever earlier than O and later than M, it will always be, and has always been, earlier than O and later than M, since the relations of earlier and later are permanent." (McTaggart 1908, 96); "every term is eternal, timeless, and immutable; the relations it may have to parts of time are equally immutable" (Russell 1903, 471).

[32] We have already quoted Smart on this in Chap. 2: "Events do not come into existence, they occur or happen. 'To happen' is not at all equivalent to 'to come into existence'." (Smart 1949, 486).

clock reading *t* with which it coincides" (215). But it is quite another to claim that "events happen tenselessly", as Grünbaum alleges Minkowski to have asserted (215). It seems to me that this whole notion of "tenseless occurrence" is a *contradictio in adjectivo*. An event occurs, happens or becomes, exactly when it occurs, happens or becomes, independently of any minds or clocks. If we say an event OCCURS, using the verb 'occurs' tenselessly, here 'tenselessly' denotes the way we have used the verb, not a variant kind of existence or occurrence. A tensed use of a verb gives implicit information about the time of utterance, just as Russell had claimed; a tenseless use does not.

We may therefore take the valid core of Grünbaum's intuition to consist in this: (i) events occur quite independently of coming into anyone's awareness of them; and (ii) one can represent an event as occurring at a certain location in the manifold without any implicit reference to the 'now' at which the event is being represented. But this in no way validates a distinction between two types of occurrence of events, "tenseless occurrence" and "tensed occurrence". An event (*eventum*, the past participle of *evenire*, Latin for to come about or happen) is something that *has become*, both semantically and etymologically. An event cannot occur without having become, since this would be to say that it could have become without having become, an evident self-contradiction. When we represent an event we therefore of necessity represent it as having become. Once we have represented all events and all processes on a spacetime diagram, we have represented all becoming, so it is unreasonable to look for something else to be superadded.

To sum up: the word 'exists' can be used temporally in a sense appropriate to things existing at a time or through time. In this sense, all events can indeed be said to be equally real, i.e. as occurring (i.e. becoming) at the particular times or spatiotemporal locations they do independently of anyone's awareness. But the spacetime manifold itself does not exist in this temporal sense. The word 'exists' or 'is' can also be used atemporally, as when we say that "event *a* IS before event *b*", and events "*a* and *b* ARE contained in the manifold". But this atemporal 'IS' is inadequate to ground any notion of events *already existing*, which clearly requires a temporal sense of 'exists'. There is, therefore, no sense of 'exists' which will support the argument that events do not need to come into existence since they (and the spatiotemporal relations among them) "already" exist in a four-dimensional manifold.[33]

Returning to our discussion of the block universe, we have seen that it is a mistake to suppose that a four-dimensional object like spacetime has an existence in time. But the mistake does not consist merely in interpreting the word 'exists' as 'exists now', as is sometimes said—it runs deeper. The manifold can neither be said to exist now, nor to remain the same through time. In fact *it does not exist at any time*. Smart makes the same point about the Minkowski spacetime of Special Relativity: "And if

---

[33] I take the view that if 'is' or 'exists' is being used atemporally, then it is a confusion to add a temporal qualification such as "at time *t*", a qualification which only makes sense for a temporal sense of 'exists'. Steve Savitt (personal correspondence) reports that this was the view of A. N. Prior regarding verbs used tenselessly, such as saying an event IS to take place tomorrow: "What place can a word like 'tomorrow' have in a strictly tenseless form?" Savitt himself allows such temporal qualification of tenseless verbs.

there can be no change in space-time, neither can there be any staying the same. As Schlick points out, it is an error to claim that the Minkowski world is static: it neither changes nor stays the same" (Smart 1964, 13). Being four-dimensional, with time included as one of these dimensions, it simply does not have a temporal existence.

This same criticism, however, also invalidates the "growing block" model, the second sense of the block universe that we mentioned at the start of this section. If spacetime as a whole does not change, then it makes no sense to think of it as growing in time either. There is no time outside of spacetime with respect to which it can grow; worse, to think of spacetime itself as a changing block is to think of it as a quasi-spatial object that can change its properties or increase in size with time.

Still, one might insist, if passage is real, there must be some sense in which this idea of a dynamism or growth is true. C. D. Broad tried to give voice to it in his *Scientific Thought* of 1923:

> Nothing has happened to the present by becoming past except that fresh slices of existence have been added to the total history of the world. The past is thus as real as the present. On the other hand, the essence of a present event is, not that it precedes future events, but that there is quite literally *nothing* to which it has the relation of precedence. The sum total of existence is always increasing, and it is this which gives the time-series a sense [i.e. direction] as well as an order. A moment t is later than a moment t^* if the sum total of existence at t includes the sum total of existence t^* together with something more. (Broad 1923, 66–7)

Thus the events that I regard as past today are more numerous than those I regarded as past yesterday; but the future has not happened yet, and therefore does not now exist. Temporal passage, on this conception, is a continual accretion of actual events.[34] Broad presented this view as clearly superior to a rival view he dubbed "The Moving Spotlight Theory" which regards "the history of the world as existing eternally in a certain order of events", and the present as a kind of moving spotlight illuminating successive events as present (1923, 59). That view, as we have seen above, illicitly appeals to a motion of something along the time dimension (here, a motion of the spotlight illuminating successive events as now), as well as regarding events as existing before they have occurred. The "growing block" view is superior in that it no longer regards events as having the kind of eternal existence imagined by Russell and McTaggart. But it still confuses the two senses of existence that I have tried to clarify above. All events we can represent in spacetime are "equally real" in that they are represented as existing in that model. However, it is only with respect to a particular vantage point within spacetime that some set of events is in the past of that event, and thus have become, and others are in the process of becoming. So you can say that at any point in spacetime there are about to be more events in the past for anything that continues to exist beyond that point. Because I have continued to exist from yesterday through to today, the set of events in my past today is greater than the set that were in my past yesterday. This is a subtle point (like so many having to do with time!). Even though it is true that from any standpoint within the universe, the number of events in the past is about to increase, there is no absolute standpoint

---

[34] Cf. Kent Gustavsson, summarizing Broad's views in this period: "Temporal passage is the continual growth of the sum total of existence" (Gustavsson 2014).

from which it is true to say that the total number of events in the universe as a whole is growing. Broad, to his great credit, appears to have come to such a realization, for he abandons this "growing block" model in his later years, as we shall see in a moment.

Such conceptions of a growing past, however, are not just the preserve of philosophers like Broad. A "growing block" model of time flow has also recently been advocated by the physicists George Ellis and Richard Muller, apparently independently of one another. Although their proposals differ in detail, they both propose a spacetime that increases by the accretion of ever more events, with the 'now' constituting a boundary or "leading edge of time" in which new events are created.[35] J. W. Dunne had criticized Bergson for holding a similar conception. For Bergson conceived his "pure duration" as the creative aspect of time, whose moments are "superposed" in a flow which creates a growing past in the face of a perfectly empty future. Dunne interpreted this as committing Bergson to "the existence of that Time embracing Time which insists on obtruding itself into every attempt at temporal analysis. His growing 'past' takes Time to grow" (Dunne 1938, 155).[36] Whether or not this is a fair criticism of Bergson, it would certainly seem to apply to the conceptions of Ellis and Muller, notwithstanding the sophistication of their physics.

But what about the expansion of the universe? Isn't my assertion about the incoherency of the growing block contradicted by this salient fact of modern cosmology? Actually, no. The expansion of the universe, discovered shortly after Dunne wrote the above criticisms, has a dual basis. There is the observational basis: as Hubble demonstrated, all the galaxies show a red shift in their spectra that is proportional to their distance away from one another. This indicates that each is receding from the other. This is then interpreted (in hindsight) as confirming what had previously been a puzzling implication of the Friedmann-Lemaître solutions to Einstein's General Relativity theory, which seemed to indicate that the metric of spacetime itself would be dynamic and not static. Combining the theory with the observation, it is concluded that the galaxies are moving apart because the geodesics defining straightline motions in spacetime for each of these galaxies are themselves receding from one another

---

[35] Ellis proposes an "Evolving Block Universe", "a spacetime which grows and incorporates ever more events, 'concretizing' as time evolves along each world line" (2014, 5). He conceives the present time as "that instant along our world line where at this moment the indefiniteness of the future changes to the definiteness of the past. It continually moves to the future, incorporating ever more spacetime events as time passes" (5). Similarly, Muller claims that "the explosion of the universe continuously creates not only new space but new time. The forefront expanding edge of time is what we refer to as *now*, and the flow of time is the continual creation of new *nows*" (Muller 2016, 10). "The future does not yet exist…; it is being created. … Every moment, the universe gets a little bigger, and there is a little more time, and it is this leading edge of time that we refer to as *now*. … *Now* is at the boundary, the shock front, the new time that is coming from nothing, the leading edge of time" (2016, 293). I am indebted to David Wright for bringing Muller's work to my attention, and for discussing it with me in correspondence.

[36] Barry Allen takes exception to this criticism. He contends that Bergson "does not even implicitly assume a universal now", so that there is nothing in Bergson's views on becoming that would commit him to the view "that becoming is not local, or that it is a single simultaneous wave front rolling across the universe" (private communication).

with a velocity that depends linearly on their distance apart (this is the dynamic metric; all this to be further explained in Chap. 7.). Still, the sense in which this is an expansion of spacetime is that each point in space is at every instant receding from every other at a certain rate, not that spacetime is an object which is getting bigger in time, as Ellis and Muller allege. That would be to conceive spacetime as though it were a purely spatial volume, and one whose size would increase in a time outside spacetime. And it would be an additional error to think of its increasing in size by the accretion of new events, something not at all vouchsafed by Hubble expansion. In sum, it is just as much a mistake to think of a four-dimensional universe as changing as it is to think of it as unchanging.

To summarize the argument so far: we have considered and rejected two versions of the block universe: (1) a conception of it as containing all events, which simply exist and therefore do not need to "become"; and (2) the growing block model, where the future is empty, and the past is growing as though spacetime were a container being filled with events in some übertime. But we can still conceive of spacetime as a block in an unobjectionable sense. This is sense (3) of the block universe. On this conception we can represent all the events of spacetime as though they have all happened, but this does not alter the fact that in order for them to have happened there must have been passage. Each event, as argued above, happens or becomes exactly when it occurs. As Dennis Dieks has expressed this view, "processes of becoming are nothing but the successive happenings of events, and … this happening of events consists entirely in the occurring of these events at their own spacetime locations" (Dieks 2006, 157).[37]

This was the considered view of C. D. Broad in his later work. He called it "absolute becoming", a notion that he introduced in a chapter of his *Examination of McTaggart's Philosophy* (vol. 2) as follows:

> To 'become present' is, in fact, just to 'become' in an absolute sense; i.e., to 'come to pass' in the Biblical phraseology, or, most simply, to 'happen'. Sentences like 'This water became hot' or 'This noise became louder' record facts of qualitative change. Sentences like 'This event became present' record facts of 'absolute becoming'. ("Ostensible Temporality", Broad 1938, Ch. XXXV, 280)

The idea is that becoming is not some extra kind of property a pre-existing event might take on, such as the property of being 'now' or entering someone's consciousness. Rather, it is constitutive of the event as an event.

Now it may be contended that this "absolute becoming" is really what advocates of the "theory of the manifold" (sense (1) of the block) were trying to get at with their notions of events "coming into being apart from anyone's awareness of them",

---

[37] "All actual events, experiences and intuitions must be there in the block representation, exactly at the spacetime position where they actually occur. … More generally, since all actual events in the history of the universe are faithfully represented, with all their characteristics and mutual relations, there cannot be anything missing in the four-dimensional picture at all." (Dieks 2006, 169). Tim Maudlin has expressed a similar view about the compatibility of the block and real temporal passage: "Insofar as belief in the reality of the past and the future constitutes a belief in a 'block universe', I believe in a block universe. But I also believe that time passes, and see no contradiction or tension between these views." (Maudlin 2002, 260).

and their rejection of accounts of passage involving motion. In a provocative article, Steven Savitt has drawn attention to such common ground between senses (1) and (3), despite the many differences already noted above (Savitt 2002). He observes that D. C. Williams—despite his championing of "the theory of the manifold" and the static theory of change, and his disparagement of the reality of passage—wants to preserve a notion of "true and real passage" in distinction from the unsound notions we have rejected here. Savitt quotes Williams' claims in his classic "The Myth of Passage" that "There is passage, but it is nothing extra. It is the mere happening of things, their existence strung along in the manifold" (Williams 1951, 463), and that the "theory of the manifold provides the true and literal description of what the enthusiastic metaphors of passage have deceptively garbled" (466; Savitt 2002, 157). Likewise, in the passage quoted at the head of this chapter, Williams writes: "World history consists of actual concrete happenings in a temporal sequence; it is not necessary or possible that happening should happen to them all over again" (Williams 1951, 464). All this sounds very like Broad's "absolute becoming", leading Savitt to exclaim: "as far as I can see, *there is no difference whatsoever between his understanding of absolute becoming and Williams' true and literal becoming*" (Savitt 2002, 160).

This is not to deny that there are some very real differences between Broad's intentions and much of what Williams writes in his "Myth of Passage", as Savitt explains. Above all, there is the difference that Williams' "true and literal description" of passage depicts it as mere temporal difference. If events are "strung along" through time "as a fence proceeds across a farm" (in Williams' colourful prose), then there is no *transition* from one event to the other. As was argued in Chap. 2, to deny passage from one point-event to another is to deny the reality of motion. But Broad's discussion in his "Ostensible Temporality" leaves his account open to the same objection. For there he begins his description of absolute becoming quoted above by claiming that "a literally instantaneous event-particle can significantly be said to 'become present'; and, indeed, in the strict sense of 'present' only instantaneous event-particles can be said to 'become present'" (Broad 1938, 280). On such a view there is becoming at every spacetime point, but there is no ground for saying that one "event-particle" comes out of another; if becoming is located only in a point, there is nothing "happening" at such a point that marks it out from what Grünbaum wants to designate as mere occurrence. Here we should remind ourselves of the origins of Grünbaum's "becomingless" view in the argument from the denseness of the continuum: if all events are point-events, there is no immediately preceding event for any event to come out of. To this we responded that if a point-event (Broad's "event-particle") is *not* taken to be ontologically primitive, as in Grünbaum's view, but is instead taken as the end-point of some process, then one can cogently maintain that something has become at any instant, even if, strictly speaking, no becoming can take place in an instant. To depict the idea that something is happening at a given point, it needs to be part of an ongoing process, and that necessitates some reference to what is going on before or after that point.

There is some evidence that Broad came to appreciate this too. In the reply to critics he gave in (Broad 1959), he describes absolute becoming as manifesting itself in "the continual *supersession* of what was the latest phase by a new phase, which will in turn be superseded by another new one" (1959, 766). If the Moving Spotlight theory is objectionable in its supposition that all events somehow "co-exist" with one another, the Growing Block is likewise at fault for presupposing "that phases, which have already supervened and been superseded, in some sense 'co-exist' with each other and with that which is now happening" (767)—that is, that past events are somehow "still there" as new ones are added. In Broad's final version of absolute becoming, the continual supersession of each phase by a new one (thus giving a directionality to process) is regarded as "the rock-bottom peculiarity of time, distinguishing *temporal sequence* from all other instances of one-dimensional order" (767).[38] In conclusion: If processes and motions are indeed represented in the block, then so must be their directionality from an initial to a final state. To deny transition, as argued in Chap. 2, is to deny the reality of motion.

3.5 Subjective Becoming

But where does the observer fit into this four-dimensional world? Surely if no account is given of where *we* are in time, namely, in the present, then something is missing from the block universe view, on any of its variants? This kind of worry, of course, is part of what motivated the critiques of mathematical time by the likes of Bergson and Whitehead that we noted at the beginning of this chapter, as well as attempts, such as those of Grünbaum and Smart, to give an account of our subjective experience of becoming without assuming the motion of a now. Let us turn to the latter first.

We have already encountered the "becomingless" view of Smart and Grünbaum, according to which events simply "occur" or "are at" the times of their occurrence, without any need for a passage from one to another. How, then, do they explain the experience of passage? The short answer is: as an experience, but only an experience. As we have seen, they distinguish sharply between the occurrence of an event, which they call "tenseless occurrence"—being at a particular time—and the coming-to-be of such an event, which they regard as a coming into the experience of an observer.

That is, they claim that although the "tenseless occurrence" they ascribe to point-events is perfectly objective, both becoming and the 'now' necessarily make reference to a perceiving subject. In this they are indebted to Bertrand Russell's distinction in his (1915) between *mental time* and *physical time*. The former, he claimed, "arises through relations of subject and object" given by time and memory, whereas the latter "arises through relations of object and object", consisting in simultaneity and

---

[38] This is similar to Whitehead's notion of the "passage of nature": "The passage of the cause into the effect is the cumulative character of time. The irreversibility of time depends on this character." (Whitehead 1920, 237). But as Hurley notes (Hurley 1986, 96), there is an idealistic underpinning in Whitehead's thinking about time indicated by such statements as "Nature is nothing else than the deliverance of sense awareness" (1920, 185).

succession (Russell 1915, 212).[39] But as McTaggart sagely noted, on Russell's view it follows that there would be no passage from past to future if there were no beings to be conscious of it:

> past, present, and future do not belong to time per se, but only in relation to a knowing subject. An assertion that *N* is present means only that it is simultaneous with that assertion, an assertion that is past or future means that it is earlier than that assertion. Thus it is only past, present, or future, in relation to some assertion. If there were no consciousness, there would be events which are earlier and later than others, but nothing would be in any sense past, present, or future. (McTaggart 1927, 13–14)

Grünbaum grasps this nettle with both hands, claiming that the becoming of an event is simply its coming into someone's conscious awareness. But then he is immediately confronted by the same difficulty we noted in Chap. 2 in connection with Barbour's Platonia: how, if there is no passage from one event to another, can there be a passage from the event of one's not being conscious of something to the event of one's being conscious of it? If there is no becoming, how can there be a "coming into awareness"?

The situation with regard to the indexicality of the now, tense, and experience, seems rather to be as follows. Whether a given event *a* occurs or not has nothing to do with whether it is registered in anyone's awareness. Grünbaum claims that the latter statement is true only if the word 'occurs' here is used tenselessly. But we can express that same statement using a tensed verb. If *a* is (tenselessly) earlier than my utterance of some statement, then it will be equally true to say that the event *has become* (by the time of my uttering this) independently of anyone's experiencing it; and if I utter this before *a*, I would say that the event *a*, if it occurs, *will occur later* than my utterance independently of anyone's experiencing it. Thus Grünbaum's claim that whereas "events happen tenselessly" (Grünbaum 1971, 215) "becoming … requires the occurrence of certain *conceptualized conscious experiences* of the occurrence of physical events" (197) seems unfounded. It is true that the 'now' does contain an implicit reference to something extrinsic to the event, but consciousness has nothing to do with it.

So if I am present when an event occurs, I will say it is happening *now*, meaning, "at the time of my utterance". But the time of my utterance is something entirely objective, even if the utterance involves me as subject. It will usually be the case, of course, that the time of my utterance is no business of physics—but not necessarily so. For example, a standard unit of measurement in cosmology is 'bya', standing for *b*illions of *y*ears *a*go, meaning billions of years before *now*, before the present era. Earth is thought to have formed, for instance, not long after 5 bya, and the Big Bang is thought to have occurred about 14 bya. As a result, textbooks on cosmology abound with diagrams where *now* is clearly labelled, contrary to oft-repeated claims that it has no place in physics.[40] Granted, this is not an instantaneous now; but the

---

[39] Grünbaum acknowledges this debt, allowing that this distinction probably has its origin Bertrand Russell's claim that "past, present and future arise from time-relations of subject and object, while earlier and later arise from time-relations of object and object" (Grünbaum 1971, 215–216).

[40] This denial of the now in physics can be quite ironic. In his *About Time* (1995), Paul Davies subscribes to the block universe view (in version 1), which depicts the 'now' as purely subjective.

point remains that the 'now' in question is indexical, picking out the time at which we intelligent observers are constructing these theories of origin. But it is also perfectly objective: the occurrence of the Big Bang 14 billion years before now is a prediction of current theories which is empirically falsifiable.

Grünbaum's position is also vulnerable to the criticism that it involves a naïve account of the subjective experience of time, where conscious experience is construed as *punctual*. In order for each point event to be experienced at a different time, Grünbaum seems, like Wells and Weyl, to be conceiving conscious experience as occupying a point of time. This conception of consciousness has been comprehensively critiqued by Daniel Dennett, who disparages it as the "Cartesian Theatre" conception (Dennett 1991, 101–138). The perceiving subject is conceived as contained in a point, like the soul that Descartes conceived as located at a point in the pineal gland in the brain, receiving the input from the nerves and interpreting it as perceptions. But as Dennett shows by analysis of some ingenious experiments by Benjamin Libet on "backwards referral in time", the reality is that the order in which events are perceived is constructed after the fact in the same process by which conscious awareness itself is constructed (139–170).

As we have seen, the idea of time as presupposing consciousness is one of the main positive contentions of James, Bergson, and Whitehead. Whitehead claimed that "actual occasions" are constituted in experience, and physics can only treat them as forming a succession in time after they have been so constituted. "In every act of becoming," he adds, "there is the becoming of something with temporal extension; but that act itself is not extensive, in the sense that it is divisible into earlier and later acts of becoming…" (Whitehead 1930, 107).[41] This echoes Bergson's claim that physics treats only the extensive aspect of time, and that this is parasitic on the creative force that is "lived time", time as experienced. Writes Bergson: "we cannot conceive a time without imagining it as conceived and lived. Duration therefore implies consciousness; and we place consciousness at the heart of things for the very reason that we credit them with a time that endures" (Bergson 1965, 48–49). Similarly, for Whitehead genetic relationships among events are only disclosed after their "concrescence" in an act of becoming, which is accounted for by a theory of their origins in "feelings of causal efficacy".[42]

But the experiments Dennett analyses tell against this idea that succession itself presupposes consciousness. When the subjects were asked to observe lights flashing consecutively and very rapidly, they often reported the flashing of the lights as occurring in the wrong order. In other words, the order in which the events ("acts of becoming") occurred successively is perfectly objective, and is presupposed in these

---

Yet he provides diagrams on pp. 133, 221, and 260 on which the 'now' is explicitly marked! (It has to be, of course, in order to discuss whether the universe began 14 bya.).

[41] Cf. also "This genetic passage from phase to phase is not in physical time" (Whitehead 1930, 434).

[42] Whitehead's ideas about the relation of time to consciousness are very convoluted; relations of succession are ultimately rooted in "feelings of causal efficacy", but the extensive aspect of such relations treated in physics requires the existence of fully conscious beings. See (Whitehead 1930, esp. 373–390); for analysis, see (Hurley 1986, 102–3).

experiments, even though the observers are conscious of them in a different order. It is only their construal of the order of events in consciousness that presupposes conscious experience, not the actual order in which the events occurred. So for all that Bergson and Whitehead wanted to situate their philosophies as monist alternatives to materialism and idealism, their constituting of duration and becoming in terms of items of (human) experience undermines that pretension, just as it does Hegel's and Russell's monist philosophies. For as soon as we construe time as presupposing consciousness, we are inexorably led into idealism, as Borges recognized: "If time is a mental process," he asks, "how can it be shared by thousands of men, or even two different men?"[43] We will return to the relation of time to experience in Chap. 6 in connection with the problem of the present in relativity.

There is one point, however, on which Whitehead and Bergson seem to me to be entirely correct. This is on the abstract character of point-events. Grünbaum wanted to take these as ontologically basic, as we saw in Chap. 2, reasoning that since there is no becoming in an instant, events are "becomingless". But as Whitehead observes, all our stock examples of events are essentially processes: the occurrence of a sea battle, the death of Queen Anne, the setting of the Sun, the emission of a beta particle from a radioactive atom. Even if these events are not all as long-lasting as the reconstruction of St. Paul's Cathedral, they all take some time. Our idea of a point-event is clearly either an abstraction from taking events like this of shorter and shorter durations, or some arbitrarily chosen terminus of such a process. Whitehead himself, using his "method of extensive abstraction", defines what he calls "event-particles" as "the ideal minimum limits to events": "points which are thus arrived at represent the ideal of events without any extension, though there are in fact no such entities as these ideal events" (Whitehead 1920, 86). Similarly he defines a *moment* as "a limit to which we approach as we confine attention to durations of minimum extension", it "has no temporal extension, and is in this respect to be contrasted with a duration which has such extension." (57) A moment, "all nature at an instant" (57), is something merely ideal, since "in truth there is no nature at an instant and there is only the abstractive set" (61); even so, this concept of "an instant of time without temporal extension … is fundamental in the expression of physical science" (61).[44]

One important emendation we should make to Whitehead's analysis of events, however, follows from the considerations of the end of the previous chapter. For him these event-particles and instants are constituents of spacetime and physical time, and as such are "spatialized" and bear no traces of the passage of time. "The state of nature 'at a moment", he writes, "has evidently lost this ultimate quality of passage" (65). We may agree with him that there is no actual becoming in an instant. Nevertheless, insofar as momentary events are derived by a limiting process from

---

[43] "Si el tiempo es un proceso mental, ¿cómo lo pueden compartir miles de hombres, o aun dos hombres distintos?" (Borges 1997, I, 418).

[44] Weyl makes a similar point: "The point-nature of the Now within the time continuum raises a certain difficulty within the conception of metaphysical time, for within a continuum, a point, without the neighborhood through which it is bound to the whole continuum, is not capable of existence. A point in a continuum is not an element of a set, but rather an ideal boundary of continuous partitions." (Weyl 2009, 32–33).

enduring processes, and thus stand for processes of arbitrary short duration, they have an intrinsic directionality deriving from processes, each of which involves a passage from its earlier states to the later. This is entirely analogous to the case with motion: just as there is no motion in an instant, and instantaneous motion is an abstraction, there is still a temporal directionality in instantaneous states and in time derivatives such as velocities, since these concepts are defined in terms of limits of changes of quantities from an initial to a final state. The importance of this consideration will become clear in the next chapter.

3.6 Summary

- In this chapter I examined the issue of the spatialization of time, arising from the portrayal of time as a fourth dimension of spacetime. I argued that the attempt to construe becoming in terms of a moving now or point of consciousness moving along a worldline is not only incoherent, but in conflict with the phenomenology of experience.
- I argued that McTaggart's conception of change in terms of changing relations to the 'now' must be rejected for analogous reasons. I argued that both this "A theory" and Russell's rival "static" or "B theory" of time are deficient in presupposing a time in which temporal relations must either change or stay the same, and "terms" that are "eternal". Similar objections apply to the "block universe" or eternalist view, according to which all the events in spacetime exist without needing to become; such a conception has more in common with Hegelian idealism than might have been supposed: it amounts to conceiving events as objects of thought, rather than as individual particulars that need to happen in order to exist.
- Another version of the block universe view, the "growing block" conception, was considered and rejected on the grounds that spacetime cannot be conceived as an object which is getting bigger in time without presupposing a time in addition to spacetime.
- I argued that point events, like instantaneous states, are abstractions, though no less vital components of our models for all that; but that insofar as they are modelling processes of arbitrary short duration, they have an intrinsic directionality deriving from processes, which involve a passage from their earlier states to the later. A view of process as involving no such transition is susceptible to the criticism that it cannot account for a crucial component of time, namely the dynamical aspect of becoming that grounds the reality of temporal succession.
- Finally I have argued that the indexicality of terms like 'now' and 'here' is no argument for their subjectivity, but only for their relational nature, and that there is no reason to expect such indexical terms to feature in scientific laws rather than in their application. All these conclusions will be important for my arguments in the following chapters.

References

Abbott, Edwin A. 1884. *Flatland*. London: Seeley & Co., Ltd.

Barbour, Julian B. 1999. *The End of Time: The Next Revolution in Physics*. (2001 reprinting). Oxford: Oxford University Press.

Bergson, Henri. 1910. *Time and Freewill: An Essay on the Immediate Data of Consciousness.* (English translation of *Essai sur les données immédiates de la conscience*, 1889). Allen & Unwin; Dover 2001.

Bergson, Henri. 1913. *Matter and Memory*. London: George Allen & Co., Ltd.

Bergson, Henri. 1944. *Creative Evolution*. (English translation of *L'Évolution créatrice*, 1907). New York: Henry Holt and Company, originally published 1911.

Bergson, Henri. 1965. *Duration and Simultaneity*. English translation of *Durée et simultaneité* (1922) by Leon Jacobson, with an introduction by Herbert Dingle. Indianapolis: Bobbs Merrill.

Borges, Jorge Luis. 1997. *Obras Completas*. Buenos Aires: Emecé.

Broad, Charlie Dunbar. 1923. *Scientific Thought*. London: Routledge and Kegan Paul.

Broad, Charlie Dunbar. 1938. *Examination of McTaggart's Philosophy*, vol. 2. London: Cambridge University Press.

Broad, Charlie Dunbar. 1959. *The Philosophy of C. D. Broad*. In ed. P. A. Schilpp. New York: Tudor Publishing Company.

Callender, Craig. 2000. Shedding light on time. *Philosophy of Science* 67 (3): S587–S599.

d'Abro, A. 1950. *The Evolution of Scientific Thought: From Newton to Einstein*. New York: Dover.

Davies, Paul C.W. 1995. *About Time*. New York: Touchstone.

Dennett, Daniel. 1991. *Consciousness Explained*. Boston/Toronto/London: Little, Brown and Company.

Dieks, Dennis. 2006. Becoming, Relativity and Locality. In *The Ontology of Spacetime*, ed. Dennis Dieks, 157–176. Amsterdam: Elsevier.

Dunne, J.W. 1938. *An Experiment with Time*, 4th ed; originally published 1927. New York: The Macmillan Co.

Dwyer, Larry. 1975. Time Travel and Changing the Past. *Philosophical Studies* 27: 341–350.

Ellis, George F.R. 2014. The Evolving Block Universe and the Meshing Together of Times. *Annals of the New York Academy of Sciences* 1326(1): 26–41. https://arxiv.org/pdf/1407.7243.

Grünbaum, Adolf. 1971. The Meaning of Time. In *Basic Issues in the Philosophy of Time*, ed. Eugene Freeman and Wilfrid Sellars, 195–228. La Salle: Open Court.

Gustavsson, Kent. 2014. Charlie Dunbar Broad. In *The Stanford Encyclopedia of Philosophy*, Spring 2017 ed, ed. Edward N. Zalta. https://plato.stanford.edu/archives/spr2017/entries/broad/.

Hawking, Stephen W. 1988. *A Brief History of Time*. London: Bantam Press.

Hinton, C. Howard. 1884. *What is the Fourth Dimension?*, reprinted in (Wells 2001).

Hinton, C. Howard. 1887. *What is the Fourth Dimension?* London: George Allen & Co, 1912 reprint.

Hurley, Patrick. 1986. Time in the Earlier and Later Whitehead. In *Physics and the Ultimate Significance of Time*, ed. David Ray Griffin, 87–109. Albany: State University of New York Press.

McTaggart, J.M.E. 1908. The Unreality of Time. *Mind* 17: 457–474.

McTaggart, J.M.E. 1927. *The Nature of Existence*. In ed. Broad C. D. Cambridge: Cambridge University Press.

Maudlin, Tim. 2002. Remarks on the Passing of Time. *Proceedings of the Aristotelian Society*, New Series, 102; 259–274.

Mellor, D.H. 1988. *Real Time*. Cambridge: Cambridge University Press.

Muller, Richard A. 2016. *Now: The Physics of Time*. New York: W. W. Norton.

Nerlich, Graham. 1994. *The Shape of Space*. Cambridge: Cambridge University Press.

Newcomb, Simon. 1894. Extract from *Modern Mathematical Thought*, reprinted in (Wells 2001).

Norton, John D. 2010. Time Really Passes. *Humana Mente: Journal of Philosophical Studies* 13: 23–34.

Pooley, Oliver. 2013. Relativity, the Open Future, and the Passage of Time. *Proceedings of the Aristotelian Society* 113 (3): 321–363.

Prior, Arthur N. 1968. *Papers on Time and Tense*. Oxford: Oxford University Press.

Russell, Bertrand. 1903. *Principles of Mathematics*. New York: W. W. Norton & Co.

Russell, Bertrand. 1912. The Philosophy of Bergson. *The Monist* 22: 321–347.

Russell, Bertrand. 1915. On the Experience of Time. *The Monist* 25: 212–233.

Russell, Bertrand. 1959. *My Philosophical Development*. London: Unwin.

Savitt, Steven F. 2002. On Absolute Becoming and the Myth of Passage. In *Time, Reality & Experience*, ed. Craig Callender, 153–167. Cambridge: Cambridge University Press.

Savitt, Steven F. 2006. *Presentism and Eternalism in Perspective*. In ed. Dieks, 111–127.

Smart, J.J.C. 1949. The River of Time. *Mind* 58 (232): 438–494.

Smart, J.J.C. 1964. *Problems of Space and Time*. New York: Macmillan.

Smart, J.J.C. 1968. *Between Science and Philosophy*. New York: Random House.

Smolin, Lee. 2013. *Time Reborn: From the Crisis in Physics to the Future of the Universe*. Toronto: Alfred A. Knopf.

Wells, H.G. 1987. In *The Definitive Time Machine: A Critical Edition of H. G. Wells's Scientific Romance*, ed. Harry M. Geduld. Bloomington: Indiana University Press.

Wells, H.G. 2001. *The Time Machine: An Invention*, ed. Nicholas Ruddick. Peterborough, Ont.: Broadview Press.

Weyl, Hermann. 1934. *Mind and Nature*. Philadelphia: University of Pennsylvania.

Weyl, Hermann. 1949. *Philosophy of Mathematics and Natural Science*. Princeton: Princeton University Press.

Weyl, Hermann. 2009. *Mind and Nature: Selected Writings on Philosophy, Mathematics, and Physics.* (Includes Weyl's 1927 essay "Time Relations in the Cosmos, Proper Time, Lived Time, and Metaphysical Time", as well as Weyl 1934). Edited and with an introduction by Peter Pesic. Princeton/Oxford: Princeton University Press.

Whitehead, Alfred North. 1920. *The Concept of Nature*. Cambridge: Cambridge University Press; reissued Mineola, NY: Dover Publications, Inc., 2004.

Whitehead, Alfred North. 1925. *Science and the Modern World*. New York: Macmillan.

Whitehead, Alfred North. 1930. *Process and Reality*. New York: Macmillan.

Williams, D.C. 1951. The Myth of Passage. *Journal of Philosophy* 48 (15): 457–472.

Chapter 4
Classical Physics and Becoming

The more closely we examine the association of entropy with 'becoming' the greater do the obstacles appear.
—Arthur Eddington, *The Nature of the Physical World* (1929, 96).

4.1 Introduction

It is often claimed that because the laws of Newtonian physics are time symmetric, there is no notion of becoming in classical physics. As a matter of historical fact, this is quite false: the notion of the evolution of the universe from some assumed initial state towards its present state, and from there towards some supposed final state was assumed by all classical physicists—not least by Newton, who worked tirelessly on Biblical chronology and who would have regarded the denial that the world was evolving towards a final state of reckoning as an abhorrent heresy. In fact, all the classical natural philosophers took for granted the reality of the order of succession, the asymmetric order in which events occur, as we discussed in Chap. 2.

But according to the majority view among modern physicists and philosophers, that is a naïve construal that misses the point. For, notwithstanding the appearance of the passage of time from past to future, Newton's laws are unable to discriminate between that direction and the future-to-past direction, so that the notion of a passage from past to future is rendered illusory at the most fundamental level of classical physics.[1] This gives rise to the problem of how to account for the fact that time only seems to "go" in one direction, from past to future, and various solutions have been offered.

---

[1] See, for example, Roberto Mangabeira Unger: "In Newton's own physics, however, no basis exists on which so to affirm the reality of time. Newton's laws of motion are time-symmetric, they supply no reason or occasion to distinguish between forward and backward temporal orderings of events" (Unger and Smolin 2015, 165). On this basis he writes of "the denial of the reality of time in this Newtonian tradition".

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_4

That, at any rate, is the received view. Let's review the arguments that have been given for it. The basic reason offered is that the equations of Newtonian physics, and for that matter, those of Maxwell's electrodynamics and Einstein's Special Relativity too, are symmetric in the time coordinate t: that is, they are unaltered when t is replaced by $-t$. So, it is inferred, there is no way to distinguish the past-to-future direction of time that we take for granted from the opposite direction, where time would go from future to past. In the words of Brian Greene,

> What all these equations have in common … is their completely symmetric treatment of past and future. Nowhere in any of these equations is there anything that distinguishes "forward" time from "backward" time. Past and future are on an equal footing. (Greene 2004, 200)

Lee Smolin offers a simple physical example to illustrate the point: the Earth's motion as it spins about its axis and orbits the Sun (Smolin 2013, 52).[2] "Reversing the direction of time reverses the direction of the orbit and the spin of the Earth, but such an orbit is also permitted by Newton's Laws" (52). The contrast is with the kinds of irreversible processes that we see all around us, like a cup falling to the floor and smashing, or milk being poured into coffee and rapidly mixing with it to form a homogeneous brown liquid. And we are all familiar with (or can easily imagine) the absurdity of the events portrayed if we film such a process and play it backwards. Such images are particularly beguiling. They have convinced most physicists and philosophers that it is irreversible processes, characterized by the increase in entropy accompanying them, that accounts for the singling out of one direction of time. From that it seems to follow inexorably that in a world where all processes were reversible, there would be no preferred time direction. On this basis it has been argued that, since all the fundamental laws of physics are time-symmetric, time is itself inherently symmetric. Thus, according to the majority opinion among philosophers and physicists, the asymmetric order in which we perceive events or states to occur is a contingent feature of the way we subjectively experience processes. In Jack Smart's words, it is a merely "epistemological asymmetry", and this is enough to explain "our subjective illusion of the one way flow of time" (Smart 1968, 288). This experienced asymmetry, on the other hand, is explained as deriving from the Second Law of Thermodynamics, according to which entropy is always increasing towards the future; the fact that we experience time as going only one way is a facet of our experience itself being an irreversible process, and it is this that grounds the direction of time.[3]

In this chapter I am going to subject these claims to careful analysis. We will begin with the conceptions of how the direction of time is related to passage offered by the two great rivals at the headwaters of modern physics, Isaac Newton and Gottfried Leibniz. We will see that although Newton's conception of time as generated by the flow of an instant is indeed susceptible to the criticisms we have already discussed (it would require a further time in which this flow can take place), there are other

---

[2] The example is taken from Lee Smolin's *Time Reborn*, (52), where he is describing the standardly accepted arguments for the unreality of time before proceeding to contest them.

[3] Similar views have been expressed by a wide range of thinkers, including Mehlberg (1961), Grünbaum (1971), Horwich (1987), Hawking (1988), Davies (1995), and Price (1996).

aspects of his view about the rate of flow that are worthy of respect. Yet Newton's model of passage is not shared by Leibniz, who characterizes becoming not in terms of any changes undergone by time itself, but instead in terms of the changing of states of an enduring thing, its present state passing into its future states. Such a view is not vulnerable to the objections to a flowing instant, as we saw in Chap. 2. On this Leibnizian model, becoming is something constitutive of process itself, whereas time is an abstraction from such processes. You would not therefore expect to derive the direction of time from considerations about time in the abstract. It will be rooted in the asymmetry of becoming of the concrete individual events and processes, and not in the laws governing types of processes.

Beginning with Leibniz, attempts have been made to base the asymmetry of the temporal ordering of events on the asymmetry of the causal order: the cause always precedes the effect. But against this it has been objected that the causal determinism of classical physicists is incompatible with any notion of objective time flow. In support of such claims appeal is made to Laplace's defining statement of the causal determinism of classical physics, which has been held to preclude not only becoming, but also contingency and freedom of the will. In Sect. 4.3 I subject these allegations to close scrutiny. It is argued that in the very statement of his position Laplace explicitly appeals to Leibniz's Principle of Sufficient Reason, according to which (in his words) "a thing cannot begin to exist without a cause that produces it". What is essential here is not the notion of cause as such, but the principle that events come to be out of prior events (the Genetic Principle). Objective becoming, that is, is presupposed in the very notion of a process in classical physics.

A second defining principle of classical determinism is the Principle of Lawfulness, according to which nothing happens in an unconditioned way. Leibniz appealed to precisely this principle in refuting the views of Newton and Clarke, who rejected determinism as being incompatible with freedom of the will. I use his arguments to show the compatibility of determinism not only with free will, but also with the contingency of events, and to throw light on the notion of causation. A cause brings about its effect through a sequence of intervening changes, and given the vagaries and complexities of the real world, such causal sequences may be interrupted by numerous accidental factors, making their connection a contingent one. So the idea of an isolated causal sequence, although necessary for the application of causal reasoning, is an idealization. The deterministic laws of physics not only abstract away from an infinity of neglected accidental factors but also, in order to be applied, require detailed knowledge of initial conditions not contained in the laws themselves.

This consideration is of the greatest relevance to attempts to define the direction of time in terms of the direction of increasing entropy. This is the subject of Sect. 4.4, where I argue that the attempt to derive the direction of time from considerations of scientific laws alone is misconceived in principle. In order to apply these laws, we need to know the initial conditions that obtain, and any asymmetry resulting from applying these laws may well be due to the initial conditions. This is relevant to Boltzmann's celebrated attempt to derive the Second Law of Thermodynamics, according to which the entropy of a given process never decreases. Boltzmann' s derivation proceeded by considering the relative probability distributions of the macroscopic states of an ideal gas constituted exhaustively by moving particles, thus characterising initial conditions solely in terms of the microstates of such particles. A persistent objection has been that since such microstates are equally probable, and the laws time-reversible, it must be impossible so to derive a time-asymmetric law like the Second Law. I use some simple examples to suggest that the asymmetry of irreversible processes derives not from the laws, but from initial conditions that are not reducible to the time-reversible microstates of particles. This objection would seem to apply even more forcefully to Boltzmann's extension of his reasoning to the whole universe, and his proposal that the direction of time is constituted by the direction of increasing entropy, at least in our region of the universe. I then discuss modern attempts to save Boltzmann's programme by appeal to the expansion of the universe, and argue that, besides objections that have been made to the hypothesis that the universe began in a state of low entropy, the very attempt to define the temporal direction of process in terms of entropy increase is misconceived in principle.

This sets the scene for a return to the claim that an asymmetric order of becoming is contradicted by the time symmetry of physical laws, the subject of Sect. 4.5. It is usually assumed that replacing t by $-t$ in the physical laws or equations governing types of processes is equivalent to reversing the direction of time—that is, to reversing the order in which the events or states making up the processes come to be. But, we shall see, transforming a given process into its time-reversed counterpart is not the same as simply having the same states A-to-B occur in reverse order of succession, B-to-A; nor even to having the time-inverses of these states B†-to-A† occur in that order. Time reversal merely concerns whether processes of type A-to-B and of type B†-to-A† occur with equal probability with respect to the standard direction of time. Thus any token of a type of process that is time-reversal invariant (i.e., that is governed by a time-reversal invariant physical law) still has its states occur in the standard time order with respect to the other processes around it.

These considerations demonstrate the fallacy of the claim described above, that if all the laws of physics were time-symmetric, the temporal order itself would be symmetric. For physical laws—whether symmetric or asymmetric in time—concern possible types of processes; and when they are applied to particular processes, a distinction between initial and final states must necessarily be presupposed. Becoming takes place at the level of individual local processes that run from initial to final states, a circumstance that is presupposed in determining whether or not a type of process occurs symmetrically in time. This is a really important point, and it is key to the position I am defending in this book. A process is something that occurs from an initial state or event to some final state or event: the order is the (local) order of becoming of the states, and this is intrinsically asymmetric.

So let us proceed more circumspectly, and take these points one by one.

4.2 Time Flow in Classical Physics

In the seventeenth century, when the foundations for modern physics were being put in place, the direction of time was a non-issue, since the one-way flow of time was taken for granted. This one-way flow was understood differently, however, by the two natural philosophers who did most to lay these foundations. In this section we are going to examine the philosophies of time of Newton and Leibniz as they pertain to this issue. What we will find is that when time is properly distinguished from the processes occurring in it, there are not two directions of time after all, but only the direction in which becoming occurs. To reverse the order in which types of states may succeed one another in time is not to reverse the order in which the actual states succeed one another in time.

When modern physicists refer to Newtonian physics, they are thinking of equations—equations such as $F = ma$ and $a = GM/r^2$, Newton's Second Law and the inverse square law for acceleration due to gravity. While these equations (together with the others constituting classical physics) are assuredly built on the foundations laid by Newton, they are the fruit of the further labour of eighteenth century thinkers such as Jacob Hermann, Jean d'Alembert, Émilie du Châtelet, Leonhard Euler, Joseph-Louis Lagrange and Pierre-Simon Laplace. Despite his seminal contributions to this algebraic approach, Newton himself had abjured such "analysis" as insufficiently rigorous, and a reader of his *Mathematical Principles of Natural Philosophy* of 1687 (henceforth, the *Principia*) will search in vain for algebraic equations.[4] He preferred proofs using geometry—albeit a synthetic geometry in which lines and figures could move, grow and have rates of increase, and in which demonstrations were achieved using proportions among such quantities. This is the kinetic approach to geometry he learned from his teacher, Isaac Barrow (1630–1677), who preceded him as Lucasian Professor at Trinity College, Cambridge. One of the great advantages of this approach, in Barrow's view, was that it avoided the difficulties stemming from composing a quantity from indivisibles. As the latter had explained in his *Geometrical Lectures* (subsequently edited by Newton before he succeeded Barrow as Lucasian Chair in 1669), a line may be conceived as "the trace of a moving point", generated in a time conceived as "the trace of a continuously sliding instant" (Barrow 1670, 165).[5] In this way, continuity is built in (as, indeed, is smoothness): at every instant of time, a point or line has a well-defined rate of growth, and geometrical figures are generated by these motions.

---

[4] I refer the reader to George Smith's excellent article in the *Stanford Encyclopedia of Philosophy*: "The modern $F = ma$ form of Newton's second law nowhere occurs in any edition of the *Principia* even though he had seen his second law formulated in this way in print during the interval between the second and third editions in Jacob Hermann's *Phoronomia* of 1716. Instead, it has the following formulation in all three editions: *A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed*" (Smith 2008, §5).

[5] As has been established by Feingold (1990, 1993), Newton would have heard Barrow's *Geometrical Lectures* (as well as his *Mathematical Lectures*) as a student in the Fall of 1664. For an analysis of the debt of Newton's ideas on fluxions to Barrow's lectures, see (Arthur 1995).

But where for Barrow the kinematic conception was a way of doing geometry that enabled it to be applied in physics, for Newton it was constitutive of the physical world. He followed Barrow in conceiving surfaces in geometry to be generated in time by moving lines, and solids by moving surfaces.[6] But he took this further. He proposed that bodies themselves would be indistinguishable from such three-dimensional geometric solids, provided God endowed them with the following three further properties: (i) being movable, (ii) obeying the accepted laws of collision, and (iii) being perceptible by the senses. His detailed argument for this account of bodies was given as a thought experiment in a lengthy critique of Descartes' natural philosophy. But although this critique lay unpublished among his papers until 1963, his idea of physical bodies being generated by motion was evident enough in his published work.[7] In a passage from *De quadratura curvarum* published in English translation in 1710, he wrote:

> I don't here consider Mathematical Quantities as composed of Parts *extreamely small*, but as *generated by a continual motion*. Lines are described, and by describing are generated, not by any apposition of Parts, but by a continual motion of Points. Surfaces are generated by the motion of Lines, Solids by the motion of Surfaces, Angles by the Rotation of their Legs, Time by a continual flux, and so in the rest. These *Geneses* are founded upon Nature, and are every Day seen in the motion of Bodies. (Newton 1710, 141)

But what are we to make of this idea of time being "generated by a continual flux" or by a "continuously sliding instant"? Taken literally, the notion seems irreparably flawed. Huw Price relates the standard objection: "if it made sense to say that time flows, then it would make sense to ask how fast it flows, which doesn't seem to be a sensible question" (Price 1996, 13). Time is not a thing that can have a velocity; and we already saw in Chap. 3 that the conception according to which its flow consists in the movement of a now from past to future is untenable. But we also saw there that there is a perfectly defensible sense of passage, where it is not time itself but any given process that advances from its state at one time to its state at another. There is no instant sliding (in what medium would it slide?),[8] but that does not rule out passage from one state at one instant to another state at another, different instant.[9] This, in fact, was Leibniz's conception of passage, to which we will now turn.

According to Leibniz, time is the order of successive things. That is to say, it is a system of temporal relations. Pick any discrete, identifiable event, such as your beginning to read this sentence. Then any other event that might occur, or have occurred, would have some temporal relation to this event: it would be either before it, simultaneous with it, or after it. This is just what time is, according to Leibniz, the

---

[6] See (Arthur 1995, esp. 341–42) for a detailed defence of the claims I make here about Newton's ontology of space, time and quantity.

[7] It was also cryptically alluded to by Locke in the second edition of his *Essay Concerning Human Understanding*, without attribution to Newton, a circumstance that intrigued Leibniz. See Dempsey's (2006) for an informative discussion.

[8] Cf. John Earman: "a literal notion of flow would presuppose a substratum with respect to which the flow takes place." (Earman 1989, 7–8).

[9] This conception of passage is also compatible with the rate of passage varying locally, as is in fact the case, as we will see in Chaps. 5, 6 and 7.

order according to which all things are situated in succession. As an order abstracted from changing things, it is not itself something subject to change.

Although time does not change, however, things certainly change in time. Without changes of state there would be no events that would be temporally related, and according to Leibniz changes in the composite things (phenomena) result from changes of state of the simple substances from which the phenomena are aggregated.[10] These simple substances are Leibniz's "monads" or units of substance. Their states are perceptions, that is, perspectival, limited, representations of the rest of the universe from their own point of view. But absolutely basic to Leibniz's ontology is the fact that each monad undergoes change of state continually, as a result of a "force striving towards change" that is the "inmost nature" of bodies.[11] It is this force, which he calls *appetition*, that "brings about change, or the passage from one perception to another" in each monad (Leibniz 1998, 269), and thus also brings about the phenomenal changes resulting from these monadic changes. The states brought about by appetition are future-directed: each state has an in-built tendency to issue in a series of states coming after it. There is, moreover, a universal harmony among states of different substances existing at the same time resulting from the fact they each represent the same universe from their own point of view. This allows Leibniz to characterize one state's being at the same instant as another if and only if they are simultaneous, i.e. represent the same universe from their own point of view, without his having to presuppose that instants exist independently. An instant is therefore effectively an equivalence class of simultaneous states, and the instants are ordered as earlier or later depending on whether the states in them are so ordered.

Without going into Leibniz's ontology in further detail, we can see that for him passage is built into the foundations of his system. Given this foundation, there is no mystery about the direction of time. Each state is future-directed: it has a tendency towards later states of the same substance, that is, those that come into existence with its passing. Even though time itself does not consist in the motion of an instant, there is passage from one state towards those later than it. The direction of time is not the direction in which some sliding instant is moving, but the direction in which states succeed one another. Thus with his depiction of time as an order of relations abstracted from processes, Leibniz is able to avoid attributing change to time, or motion to the instant generating it. If time is constituted by the order of succession, then all those events or states b that could come into existence out of event or state a will be such that *a is before b*, which we may write aBb. The relation is clearly asymmetric: if aBb then it is not the case that bBa. Any given process whose initial

---

[10] As Leibniz explains to De Volder, "You doubt, distinguished sir, whether a single simple thing would be subject to changes. But since only simple things are true things, the rest being only beings by aggregation and thus phenomena, and existing, as Democritus put it, νόμῳ [by convention] not φύσει [by nature], it is obvious that unless there is a change in the simple things, there will be no change in things at all." (GP II 252).

[11] These quotations are from Leibniz's essay *Specimen Dynamicum* published in the journal *Acta Eruditorum* in April 1695 (Leibniz 1998, 155). Cf. also "I also take it for granted that every created thing is subject to change, and therefore the created monad as well; and indeed that such change is continual in every one" (*Monadology* §10; Leibniz 1998, 269).

state or event is a, will be such that any of its other states or events, say x, will be related to a by aBx. It will have a temporal direction from a to x. Now, if all events and states, and not just those of a given process, can be ordered in such a way that each comes to be only out of events and states that are earlier than it, this asymmetrical ordering is what constitutes the direction of time.[12]

Thus Leibniz is content to define time as "the order of succession". But as Samuel Clarke objected on Newton's behalf, "space and time are quantities, which situation and order are not" (Clarke's Third Reply, Leibniz 1969, 685). For Newton, quantity is not reducible to order, but presupposes an existing measure. This was one of his reasons for believing that mathematical quantities like space and time must be presupposed as existing prior to the things in them. Leibniz, on the other hand, had his reasons for believing that you could ascribe quantity to the order of succession, as he conceived it.[13] But without getting into that rather technical issue, we can see what Newton and Clarke are getting at. For as Clarke had objected, the order among certain successive existents could remain the same, however quickly they might succeed one another.[14] How, then, do you determine the quantity of time, that is, how fast time passes?

Leibniz's answer is the traditional one: we measure time by reference to uniform motions. In one of his many unpublished manuscripts, this one written some thirty years before Clarke's challenge, he had written:

> We measure time by some uniform change applied to the parts of an enduring thing, in such a way that the earlier is that which is applied to a part less distant from the assumed starting point.
>
> The basis for measuring the duration of things is the fact that agreement is obtained if several uniform motions are assumed (such as several accurate clocks). (A VI 4, 629/Leibniz 2001, 275)

As Aristotle had said, "not only do we measure change by time, but we also measure time by change" (*Physics* iv 220 b14), so that "time is the number of change with respect to before and after" (*Physics* iv 220 a24). Aristotle speculated that the reason some of his predecessors had identified time as the uniform motion of the heavenly sphere is that "all other motions are measured by this motion, as is time too" (*Physics* iv 223 b21). And the reason they picked out the motion of the heavenly sphere—that is, from a Copernican point of view, the apparent daily rotation of all the heavenly bodies resulting from the Earth's spin—is that this was the most uniform motion

---

[12] Leibniz himself assumed that every event is either before, simultaneous with, or after, every other event, and proposed an "axiom of connectibility" to ensure this (see Arthur 1985). But the asymmetric order of events need not be such a total ordering; it can be a partial ordering, as in Minkowski spacetime (see Chap. 6).

[13] Leibniz's answer to Clarke was that "order also has its quantity; there is in it that which goes before and that which follow; there is distance or intervalv." (Fifth Paper, §54; Leibniz 1969, 706). An order of successions would be a succession of intervals, or "like states", each possessing a duration whose length could be determined by comparison with a standard.

[14] "The order of things succeeding each other in time, is not time itself: for they may succeed each other faster of slower in the same order of succession, but not in the same time." (Fourth Reply §41; Leibniz 1969, 695).

evident to them. This was not practical for everyday concerns, however, and for these purposes the Ancient Greeks had time-keeping devices that depended on other uniform motions: the clepsydra (water-clock); and later, the clepsammia (sand-glass, or hour-glass). An hour-glass depends on the flow of sand, a clepsydra on the flow of water, into a suitably calibrated vessel.[15] Although these devices were accurate enough for their purposes, it was the daily rotation of the heavenly sphere that the ancients appealed to in assessing the reliability of these clocks.[16]

The accuracy of a clock therefore amounts to the uniformity (or "equableness") of the motion or flow by which it is measured. The ancients, as we have seen, relied on the uniformity of the daily rotation of the heavenly sphere. But of course, as the days are not all of an equal length, especially in latitudes far from the equator, the year—the apparent annual motion of the Sun—yields a more accurate measure. The mean solar day would then be an exact fraction of the solar year. This was still taken for granted in the time of Pierre Gassendi (1594–1658), Descartes's main rival and source of the notion of absolute time taken up by Newton: "there is no other motion more universal, more constant and better known, than that of the Sun" (Gassendi 1649, 623). But how do we know the Sun's motion is equable? Barrow had recognized the issue in his *Geometrical Lectures*. He followed Gassendi in holding that time lapses at an even tenor, independently of anything existing in it. For him this is an absolute "Quantum". But its quantity or measure is determined by an equable flow, and we only ever have access to comparison of such flows through our timekeeping instruments. Thus to determine whether the Sun's motion is equable, our only recourse is to compare the Sun's motion with other equable motions, such as provided by the clepsydra or hour-glass. Here, though, Barrow introduces a momentous inversion: for him the "primary and original motions" are *not* the celestial ones, but rather local, mechanical ones, i.e.

> those motions which we observe by the senses and which are subject to our experiments, since by means of these we are able to judge the regularity of the celestial motions. Nor is even the Sun itself the arbiter of time or a reliable witness, except insofar as its veracity is assured by mechanical timekeepers. (Barrow 1670, 161; Čapek 1976, 206)

Here Barrow had in mind the practice of the astronomers of his time, who "very accurately" measured times (for instance those of eclipses or transits) by the use of a pendulum. But the appeal to such mechanical motions as standards of equableness

---

[15] The clepsydra (literally, "water-thief") dates back probably thousands of years, but is known to have been used in Aristotle's time to keep time on clients' visits to brothels in Athens, and increasingly sophisticated mechanisms (pointers, gears, and escapement mechanisms) were subsequently developed by the Greeks and Romans (as well as, independently, the Chinese) to help improve its accuracy. See Landels (1979, 33).

[16] Lest it be thought that the Greeks were technologically primitive, mention should be made of the Antikythera mechanism retrieved from a shipwreck off the coast of the Greek island of that name in 1902. Subsequent analysis has revealed that this was a clockwork device whose 37 gear wheels would have enabled its users to follow the movements of the heavenly bodies, sufficiently for them to be able to predict eclipses. The fact that it could even model the irregular orbit of the moon has led to speculation that the 2nd century BC astronomer Hipparchus of Rhodes might have been consulted in its construction. For a full account, see Jones (2017).

also reflects the practical situation in the seventeenth century, where the problem of accurate timekeeping was very much to the fore. Although latitude could be easily calculated from the angle of the Sun at midday, ships could only calculate their positions regarding longitude by keeping track of their own speed and how long they had been travelling. This had led to various extremely costly marine disasters, where gold-filled galleons had been wrecked on rocks which their captains had calculated were hundreds of miles away.[17] The time could be estimated by “shooting the Sun”, but the Sun’s angular position at midday does not mark off equal intervals throughout the year, leading to significant inaccuracies if the days are assumed to be equal. Astronomers had known since Ptolemy that the non-uniformity of the Sun’s apparent motion was due to the eccentricity of its annual orbit and its obliquity to the celestial equator—in Copernican terms, because the Earth orbits in an ellipse with the Sun at one focus and because its axis of rotation is tilted relative to the normal of the plane of its orbit. Ptolemy had corrected for both these irregularities by using sidereal time instead of solar time, that is, by basing the year on the time it takes a given star to make one revolution (this would also take into account the fact that, relative to the Earth, the Sun makes one revolution a year, so that the mean solar day is about 1/365 of a year or 4 min more than a sidereal day). Using sidereal time, one could calculate how the solar days at a given latitude varied in length during the year, the so-called “equation [i.e. equalisation] of time”.[18]

The development of a new, accurate pendulum clock by Christiaan Huygens in 1661 was therefore of the first importance. And, not surprisingly given the above context, one of the first applications that Huygens made of it was to provide a set of tables for the equation of time, published in his *The Pendulum Clock* in 1673 (Huygens 1986, 26–27). These tables allowed mariners to correct for the above-mentioned deviations from the sidereal day over the course of a year: having determined when the sun was directly overhead on a given day, they could then add or subtract a few minutes or more according to Huygens’s tables.

These considerations allow us to understand Newton’s position. For (as we shall see further in Chap. 5) his claim in the *Principia* that “absolute time … flows equably without relation to anything external” has been widely condemned by philosophers of an empiricist bent, like Ernst Mach. But in contrasting this “absolute, true and mathematical time” with “relative, apparent and common time”, Newton is explicit in identifying the quantity of absolute time by its *mathematical* measure, as given by the equation of time, in contrast with the *empirical* measures of supposedly uniform motions like “an hour, a day, a month, a year” associated with the motions

---

[17] For a vivid account of the problem of longitude and John Harrison’s patient (and inadequately rewarded) solution of it through the construction of clocks accurate enough to determine longitude, see Sobel’s (1995).

[18] Julian Barbour gives a pellucid and highly informative account of Ptolemy’s contribution to time measurement in his (2001), esp. pp. 175–190. As he observes, this selection of sidereal time by Ptolemy was a major contribution to the scientific revolution, in that it made it “possible to give precise content to the notion of a universal and equable flow of time” (Barbour 2001, 179); “it is only when the one sidereal time is used that all motions can be simultaneously described by simple theories. The time is therefore universal.” (2001, 180).

of the heavenly bodies. “In astronomy,” he writes in the *Principia*, “absolute time is distinguished from relative time by the equation of common time” (Newton 1999, 410). In an earlier draft he had incorporated this reference to the equation of time into his definitions of absolute and relative time:

> Def. 1. Absolute time is that which flows equably by its own nature and without relation to anything else. Such is that whose equation the Astronomers investigate, otherwise known by the name Duration.
>
> Def. 2. Time regarded as relative is that which is considered equable in respect of the flow or transition of some sensible thing. Such is the time of days, months, and other celestial periods according to the common way of thinking.[19]

Thus any time whose measure is given by a particular motion taken to be uniform (such as that of a pendulum clock or the eclipses of the moons of Jupiter) is only a *relative time*. But this raises the prospect that the time beaten out by one such regular motion, say that of the Jovian eclipses, would appear irregular when compared against the periodicity of another, say the pendulum. Barrow’s response to this situation, we saw above, was that we should simply take “mechanical timekeepers” as giving us more accurate measures of time. But this is an assumption that seems to require justification, and such a justification is what Newton sought to provide.

The basic intuition underlying Newton’s conception of time flow is this. It cannot be merely a matter of convention which processes we take as uniform, to serve as appropriate measures of the equable flow of time. There is a fact of the matter of how long a galleon has been sailing, but if this is measured by reference to the relative time given by the motion of the Sun, shipwrecks result. Barrow’s insistence on the superior accuracy of mechanical timekeepers like Huygens’s pendulum clock is therefore not arbitrary.[20] Huygens himself, of course, was no mere experimental physicist, and had in fact given a full theoretical justification for the isochrony of Galileo’s pendulum for small oscillation amplitudes, and for his own clock with its cycloidal flanges.[21] This was based explicitly on three hypotheses that Newton would take as his own starting point, the first of which is a combination of Descartes’s first two Laws of Nature as applied to falling bodies, and a version of what Newton would promote as the Law of Inertia:

> *Hypothesis I*: If there were no gravity and the air did not impede the motion of bodies, then any body would continue its given motion with uniform velocity in a straight line.

---

[19] (Herivel 1965, 304). This draft, written probably in 1685, was entitled *On the Motion of Bodies in regularly yielding media*; translation from (Arthur 1985, 348).

[20] In his *Pendulum Clock* Huygens reports an account of a sea voyage in 1664 by Alexander Bruce, who had been equipped with an early version of his clock. After sailing westward from the island of São Tomé for seven hundred miles in the company of three other ships, then turning south-southwest towards Africa for another two or three hundred miles, they were running out of drinking water. Bruce’s calculations of their position as just 30 miles off the island of Fuego of the Cape Verdi Islands (where they were able to dock the next morning) differed by 80, 100 and more than 100 miles from those calculated by the other three ships lacking his clock (Huygens 1986, 28).

[21] The full title of Huygens’s book is *The Pendulum Clock, or Geometrical Demonstrations Concerning the Motion of Pendula as Applied to Clocks* (Huygens 1986).

> *Hypothesis II*: By the action of gravity, whatever its sources, it happens that bodies are moved by a motion composed of both a uniform motion in one direction or another, and of a motion downward due to gravity.
>
> *Hypothesis III*: These two motions can be considered separately, with neither being impeded by the other. (Huygens 1986, 33)

The upshot of this is that the quantity *time* appearing in Huygens' analyses of falling bodies and of pendulum bobs taking a cycloidal path, is identical to the time corresponding to a uniform, that is, inertial, motion. This is not an empirical measure, but rather the mathematical time on which the analysis of bodies moving under the action of forces is based. Thus Huygens deserves credit not only for bringing the measure of the flow of time down to Earth,[22] as it were, but also for clearly articulating its mathematical basis. Newton is following his lead when he says that astronomers correct the inequality of the natural days "in order to measure celestial motions on the basis of a truer time" (Newton 1999, 410). But, he warns, "It is possible that there is no uniform motion by which time may have an exact measure" (410). No doubt he had in mind his theory of universal gravitation. For if every massive body attracts every other, then bodies can at best approximate inertial motions when they are as far removed as possible from ponderous matter.

To summarize: Newton follows Gassendi and Barrow in believing that time lapses at an even tenor independently of anything else. Barrow believes that this is the "Quantum" of time, whose Quantity is given by an equable flow, and that we only ever have access to a comparison of such flows through the mediation of timekeeping instruments. On his analysis, the time taken for a particular process will then be measured by some physical process taken as uniform or equable: this is what Newton calls a relative time. But if, as Barrow proposed, we simply adopt one measure as giving an equable time, namely that given by certain local motions, this appears to make the equableness of the flow of time conventional, a mere matter of convenience or simplicity.[23] Newton rejected this as empirically untenable, as well as incompatible with Barrow's assumption that these relative times are measures of a universal quantum, which he (Newton) took to be the equable flux of absolute time.

Leibniz, as we have seen, rejected the notion of time as a thing capable of flowing. But he did not reject the absoluteness of time in the sense just identified, the mathematical time assumed by Huygens as a unique parameter in the equations of motion. He could agree with Newton that this mathematical time indeed represents the uniform flow of time that is marked out by an inertial motion, and that we identify which motions are inertial by the calculation of forces. But by flow of time he meant how quickly world-states follow one another, not how quickly some instant is moving

---

[22] As Julian Barbour comments, "Huygens could be said to share with the Hellenistic astronomers the credit for the practical application of the way in which the passage of time is manifested in the world" (Barbour 2001, 455).

[23] Intriguingly, Barrow's view can be seen as an anticipation of the situation in canonical quantum gravity. According to the reading of Carlo Rovelli and others, the introduction of a Newtonian universal time parameter is incompatible with general relativity and its quantization, so that the only recourse is to define time in terms of a particular physical process. See Rovelli (2018) and Kiefer (2011).

or sliding. Now, it is perhaps a matter of convention to regard this flow as uniform. For, providing all the phenomena are "marching in step", to use Julian Barbour's happy phrase (Barbour 2008, 3–4), then any variation in the rate of flow would affect them all in the same way, and would therefore be in principle indiscernible. But the supposition that there is one such flow is not a convention. Rather, it is a constitutive assumption of classical physics that is borne out by Newton's comprehensive success in applying that mathematical model to the solar system.

The hypothesis of absolute time in this sense of one universal flow is, of course, refuted by the success of relativity theory, as we shall see in detail below. Einstein showed that the flow of time in some process P as measured by some standard clock C will in fact vary with the relative motion of C with respect to P. So while it may be the case that the duration of things *in themselves* "remains the same, whether their motions are rapid or slow or null" (Newton 1999, 410), the rate of flow of the same process will differ as measured by clocks in differing states of motion with respect to it.

Secondly, though, it is this very correlation of time with motions in space that opens up the possibility that time could flow at different rates for bodies taking different paths through spacetime, as we shall see to be the case in the Minkowski spacetime of Special Relativity. This results in the apparently paradoxical situation (the Twin Paradox, to be discussed in Chaps. 5 and 6 below) where twins taking different paths through spacetime would discover, on reconvening, that they had aged differently. The phenomena no longer march in step at each instant of an absolute time, but they still are correlated in a non-conventional way through the fact that inertial motions through spacetime embody the standard of rate of flow, thus preserving this aspect of Newton's insight. As we shall see in later chapters, this tying of a standard and non-conventional rate of flow to inertial motion is preserved in Special Relativity and in General Relativity too, albeit with an altered understanding of what inertial motion consists in.

As for Leibniz, although he took for granted the existence and privileged nature of inertial motions for determining the quantity of time, he did not sufficiently take into account that such motions require space as well as time. So his clean separation between time as the order of successive things and space as the order of simultaneous things is not adequate to ground the representation of motion in a straight line.[24] But that is not to say that Leibniz neglected the fact that there must be a spatiotemporal interval between any two events one of which is the cause of the other, and therefore (assuming the principle of retarded action) a temporal interval. As we shall see, he used this very fact as the basis of his sketches for a causal theory of time. This is of great interest since this connection between temporal and causal precedence (together with some other natural assumptions) can be used to derive Special Relativity, as will be discussed in Chap. 6. But in order to properly situate that, we should first examine the notion of causation and determinism in their classical contexts. This will also serve as a valuable context for our discussion in Chap. 8 of claims made about

---

[24] This was established with great precision by Howard Stein in his "Newtonian Space-time" (Stein 1968).

indeterminism in quantum physics, and the challenge which that theory is alleged to pose for the scope of the causal principle.

4.3 Time, Cause and Determinism

Classical physics is usually presented as entailing a rigid determinism. This is thought to follow straightforwardly from the mechanical philosophy, where the world is conceived as consisting in micro-corpuscles and their motions, governed by Newton's Laws. Thus, in the words of Paul Davies, "Newton gave us rigid determinism, a world of inert particles and forces locked in the embrace of infinitely precise lawlike principles" (Davies 1995, 31).[25] As we shall see, there is more than a little historical revisionism here. Newton was vehemently opposed to such determinism, regarding it as refuted by the ability of intelligent agents such as ourselves to initiate motions by our will alone. He believed that the natural world could *not* consist solely of inert particles following the three laws he had presented, but that there must also be active principles, including not only human and angelic agents, but also the non-mechanical powers he proposed were responsible for electricity, magnetism, gravitation, and even fermentation. Here I am not just quibbling about the inaccuracy of the usual potted histories; the point is that such accounts betray a lack of clarity about what the assumption of determinism entails, and a more accurate historical account will serve to reveal several nuances in the relationship among determinism, causation and predictability.

The classic statement of determinism in classical physics was given by Pierre-Simon de Laplace in the introduction to his book on probability theory in 1812:

> We ought, then, to envisage the present state of the universe as the effect of its earlier state and as the cause of the state that is to follow. An intelligence that could, for a given instant, know all the forces by which nature is animated and as well as the relative situations of the beings that compose it, would comprehend in the same formula the motions of the greatest bodies in the universe and those of the lightest atom, provided it were sufficiently vast to submit all the data to analysis; to it, nothing would be uncertain, and the future, like the past, would be present to its eyes. The human mind, in the perfection that it has been able to give to astronomy, affords a feeble outline of such an intelligence. (Laplace 1812, Introduction, ii–iii; my translation)

Much has been written about the implications of such determinism. Not only has it been claimed, as by Newton, that it precludes free will, but also that it entails the necessity of all that happens, making contingency only another name for human ignorance, as it was for Spinoza. It has been held to rule out genuine novelty, and even temporal succession. In the words of Milič Čapek (expounding Bergson), "'Future' is merely a label given to the unknown part of reality which *coexists* with our present moment in a sense similar to that in which distant scenery coexists with our limited

---

[25] Cf. Brian Greene: "According to Newton, if we knew in complete detail the state of the environment … we would be able to predict (given sufficient calculational prowess) with certainty whether it will rain at 4:07 p.m. tomorrow" (Greene 1999, 91).

visual field".[26] Consequently, Čapek asserts, we are left with two and only two choices: "*either* real succession with an element of real contingency, *or* complete determinism with the total absence not only of possibilities, but of succession as well" (Čapek 1971, 111).[27] As we will consider further in Chap. 8, such rigid causal determinism is often contrasted with the indeterminacy of quantum theory, which has been thought to allow a reinstatement of the real contingency necessary for free will. Thus according to Herman Bondi, "the flow of time has no significance in the logically fixed pattern demanded by deterministic theory, time being a mere coordinate. In a theory with indeterminacy, however, the passage of time transforms statistical expectations into real events".[28]

But does determinism really entail a strict necessitarian worldview that leaves no room for contingency or time's passage? Laplace is often credited with having been the first to clearly articulate such a view, on the basis of the above remarks. His image of the all-knowing and inconceivably vast intellect has been dubbed "Laplace's Demon", perhaps with Descartes's and Maxwell's rather different demons in mind. But as Laplace himself makes perfectly clear just prior to the quoted passage, the omniscient intelligence he alludes to there is simply the divine being to whom Leibniz had earlier imputed the same powers. Actual events, he claims, have a linkage with preceding ones founded on Leibniz's *principle of sufficient reason*, whereby "a thing cannot begin to exist without a cause that produces it" (Laplace 1812, ii).

Leibniz's principle, of course, was theologically based. God would create nothing without a reason, so that there would be determinate reasons for everything that happens in the world he decides to create. If an agent acts freely, God would know the motives determining the action, as well as everything else that must be in place in order for the action to be performed, and would create everything accordingly. But, Leibniz insisted, creation is not causation; in order for the actions to be ascribed to the agent and not to God, the agent would have to be equipped with the means to produce its own actions, according to a law of succession that is intrinsic to that agent. But in order for this to cohere with everything else, each enduring thing in the world (each substance) would likewise have to have a law of production of its states known beforehand to God, so that it would pass from its present state into its future states in accordance with its own law and in harmony with the states of all other substances. Thus all states are produced from preceding states in a pre-determined order of succession specific to each substance: they come to be out of those states in an

---

[26] Čapek (1971, 111). Cf. Bergson on Laplacian determinism: "Radical mechanism implies a metaphysic in which the totality of the real is postulated complete in eternity, and in which the apparent duration of things expresses merely the infirmity of a mind that cannot know everything at once" (Bergson 1944, 45).

[27] The same point of view is expressed rhetorically by G. J. Whitrow, when he asks "if the future history of the universe pre-exists logically in the present, why is it not already present? If, for the strict determinist, the future is merely "the hidden present", whence comes the illusion of temporal succession?" (Whitrow 1961, 295); quoted from (Grünbaum 1971, 226).

[28] (Bondi 1952, 660); Grünbaum reports that by 1971 Bondi no longer subscribed to this view (Grünbaum 1971, 220).

inherently asymmetric ordering. There is therefore nothing in Leibniz's conception of determinism that precludes the reality of succession or becoming: in fact, it requires them.

Laplace, on the other hand, is famous for having allegedly told Napoleon that God is "a hypothesis I do not need in my philosophy". So one might wonder how he saw himself entitled to the principle of sufficient reason, given its theological foundation. Yet Laplace could still assume that there is always a sufficient reason why an event occurs, even if it is opaque to us, without needing to appeal to a divine guarantor. All that is needed is the assumption that every event is produced in a determinate way out of previous events, even if we do not know the law of production. In fact, this is the assumption on which his entire theory of probability proceeds.

In sum, a process is simply a sequence of events produced out of prior events according to some principle or principles of determination. The order in which events come out of events in their past is the order of succession, an inherently asymmetric ordering. There is no process without becoming. But deterministic theories certainly represent processes. So the idea that there is any incompatibility between deterministic theories and the reality of succession or becoming is simply false.[29]

This principle that events come to be out of prior events, evident in both Leibniz and Laplace, is called by Mario Bunge "the *genetic principle*, or principle of productivity" (Bunge 1979, 25–26). It is one of two principles he has identified as constitutive of determinism in general. The second is the *principle of lawfulness*, according to which "nothing happens in an unconditional and altogether irregular way—in short, in a lawless, arbitrary manner" (26). This principle is also explicitly espoused by Leibniz, and again an examination of his position will help to clarify the nature of determinism. In his *Discours de metaphysique* Leibniz insists that "not only does nothing happen in the world that is absolutely irregular, but also we can't even imagine such a thing" (A VI 4, 1537/Leibniz 1969, 306). In support, he argues as follows:

> Suppose, for example that someone puts a number of completely haphazard dots on a piece of paper, as do people who practise the ridiculous art of geomancy. I say that it is possible to find a geometrical line whose notion is constant and uniform according to a certain rule, such that the line passes through all the points, and in the same order in which they were drawn. … [Likewise] there is no face whose contours are not part of a geometrical line, and which could not be drawn in a single line by some rule-governed motion. (*Discours* §6, A VI 4, 1537–38/Leibniz 1969, 306)

Of course, Leibniz grants, if the line is very complex, it will appear irregular, even if it is the result of a divinely imposed order. "Therefore what counts as extraordinary is only so with respect to some particular order established among created beings. For as regards the universal order, everything conforms to it." (A VI 4, 1538/Leibniz

---

[29] Cf. Earman, who argues "that there is a well-defined sense in which Newton's laws of motion with certain force functions, though invariant under time reversal, allow forward but not backward causation in the sense that past states affect future states but not vice versa. If true, something along this line would be sufficient to account for the asymmetry of traces with respect to past and future" (Earman 1974, 42). See also Earman's (1986) for a trenchant critique of the usual view of classical physics as deterministic, on various grounds independent of those I discuss here.

1969, 306) Here Leibniz is distinguishing lawfulness as it corresponds to known physical laws, from the lawfulness corresponding to the divine order instituted by God.

In this respect, Leibniz's position is more sophisticated than Laplace's, at least, as it has traditionally been interpreted. For Laplace assumes as a theoretical posit knowledge of all the initial conditions of the universe, but then also assumes that a given set of laws that are in principle accessible to the human mind will govern all subsequent development. Indeed, these are often assumed to be Newton's laws, as we saw in the quotation from Davies above. For Leibniz, by contrast, God's knowledge of the "laws of the series" for each existing thing, together with his omniscient grasp of initial conditions, constrains all subsequent events to be determined. But God knows such things through a divine intuition, and precisely not in the sense that one event can be demonstrated to follow logically from prior ones. For something to be logically necessary is for its opposite to involve a contradiction. But physical laws do not license this kind of necessitation: *the infinite complexity of the initial and intervening conditions preclude such a logical deduction of any existing event from another.*

What this discussion of Leibniz's position shows, then, is that one can commit to an ontological determinism, according to which all events follow from preceding events in a lawful way, without being committed to the view that whatever happens according to the laws we discover involves any logical necessitation. Not only do we not know for certain whether we have taken into account all the relevant conditions for the applicability of these laws; even the laws we have abstracted do not entail their consequences with logical necessity, but only contingently on our not having overlooked relevant factors in abstracting them from the vagaries and complexities of the actual world. So contrary to what Bergson and Čapek allege, there is no conflict between determinism and the emergence of qualitative novelty: that charge depends on a conflation of determinism—the thesis that all events are produced according to regularities describable by laws—with logical necessitation. In the same way, Whitrow's claim that "the future history of the universe pre-exists logically in the present" (Whitrow 1961, 295) is seen to be without foundation.

So far then, we have seen that there is nothing in determinism to impugn the reality of temporal succession, becoming, contingency or novelty. What about free will? It is frequently claimed that both free will and contingency are undermined by the claim that there is nothing unconditional in a deterministic universe—the content of the principle of lawfulness, as described above. Leibniz, however, appeals to this very feature in his demolition of the conceit that determinism is incompatible with freedom of the will. So it will be worth exploring his views a little further. For Leibniz conceived all events, including contingent ones such as actions of an agent with free will, to be subject to his principle of sufficient reason. Indeed, Laplace noted this with approval just prior to the "demon" passage quoted above:

> This axiom, known under the name *principle of sufficient reason*, extends to even the most indifferent actions. The freest will cannot be without a determining motive, giving birth to it; for if all the circumstances of two situations [*positions*] were exactly the same, then if the

> will acted in the one case but refrained from acting in the other, its choice would be an effect without a cause. It would be, says Leibniz, the blind chance of the Epicureans. (Laplace 1812, ii)

This correctly describes the position Leibniz had taken in his celebrated controversy with Samuel Clarke. As we have already remarked, Newton took Leibniz's conception of determinism to preclude the possibility of free will, and on this Samuel Clarke took Newton's side. Clarke claimed that Leibniz, in likening the will to "a balance where reasons and inclinations take the place of weights" (GP VII 359/Leibniz 1969, 680), was reducing the world to a mere clockwork mechanism, subject to fatal necessity. But on Clarke's own view (as on Newton's), the mind, being active, has a "self-motive principle" by means of which it can choose in the absence of determining causes. The mind, he asserts, may have reasons for acting even "when there may be no possible reason to determine one particular way of doing the thing rather than another" (Fifth Reply, §§1–20/GP VII 422). To Leibniz this seemed an obvious self-contradiction. If the mind has reasons for its choice, then there must be a possible reason for it to choose one alternative as better than the rest. The mind "acts by virtue of its motives, which are its dispositions to act" (Fifth Paper, §15, GP VII 392/Leibniz 1969, 698). To imagine that the mind can act in defiance of its own motives is "to divide the mind from its motives, as if they were outside the mind as the weight is distinct from the balance, and as if the mind had, besides its motives, other dispositions to act by virtue of which it could reject or accept the motives" (392/698). Here Leibniz seems to be perfectly correct. If there is no reason for the mind to determine to do one thing rather than another, then it does not have sufficient reason to act. It is only if the mind does have a sufficient reason for its course of action that one would want to count the act as freely chosen. So, Leibniz concludes, having a sufficient reason is required for freedom of choice, not contrary to it.

On this view, agents are free to act to the extent that they are the authors of their own actions: they must be autonomous agents. The deciding motive of the individual agent, together with all the other conditions necessary for the action, would then constitute its full cause. Since the agent's autonomous participation is itself a necessary condition for the act to take place, there is no question of the act being externally compelled just because it is fully determined. This is illustrated by the apocryphal story told about the Stoic philosopher Zeno of Citium catching his slave stealing. When the clever slave tried to argue that he was not responsible for stealing from his master because he was "fated to do it", Zeno replied, "And I am fated to flog you for it!". So we see that being determined is not the same as being caused to happen. As Leibniz sagely points out, the notion that an event's being determined means that it will happen no matter what you do, is a fallacy, "a sophism that has troubled people in almost every age" (Leibniz 1951, 54):

> It is false that the event happens whatever one may do: it will happen because one does what leads to it; and if the event is written, the cause that will make it happen is written too. Thus the connection of effects and causes, so far from establishing the doctrine of a necessity detrimental to conduct, serves to destroy it. (Leibniz 1951, 57)

Every event must have a cause, but a cause does not necessarily produce its intended effect. In order for a cause to produce its effect, other conditions must concur.

The operative notion of cause here is that the complete cause of something should be thought of as the sum of all its "requisites", that is, the conditions necessary for a thing's existence.[30] Among these, there is usually one isolable as the one that initiates the action in question, and this is what is customarily labelled the cause. In the above example, the slave's decision to steal would be called the cause. In the case of the evaporation of water on the pavement on a hot day, it would be the sun's rays. Determination, though, is not necessarily causal: not all scientific laws are causal laws. For example, the later state of a thing could be determined by its earlier state and some functional relationship between them. But a state does not initiate an action, so an initial state is not the cause of succeeding states. Not only this, but functional relationships are usually reciprocal, contrary to the unidirectionality needed for causation. In addition, there is also statistical determination, and this is relevant to the claims noted above about the indeterminism of quantum physics. For the quantum processes that exhibit indeterminism, such as the indeterminism of the trajectory of an electron going through a magnetic field, are still in conformity with the statistical determination prescribed by the theory, and it is this very fact that allows for an approximate determinism to be regained "in the large" (i.e. for large numbers of similarly prepared systems, or for processes whose action is very large compared with Planck's quantum of action).

A cause brings about its effect through a sequence of intervening changes. Leibniz exploited this idea to formulate a causal theory of time. For him, the sequence of intervening changes is what distinguishes causes (which he terms "mediate requisites"), from "immediate requisites", requisites whose existence is necessary for the thing to exist but which do not produce it (such as the parts in a whole, or points in a line). Thus "An [efficient] cause is a requisite according to that means by which the thing is produced." If two things are such that "the first is the condition of the second by an intervening change, then the first is earlier, the second later." (mid-1685; A VI 4, 628). On this basis Leibniz erects his sketches of a causal theory of time:

> If one thing is the cause of another, and they are not able to exist at the same time, the cause is *earlier*, the effect is *later*. Also earlier is whatever is simultaneous with the earlier.[31]

Something that brings about an effect through a sequence of intervening changes constitutes a *causal process*. As conceived by today's physicists, a causal process is an isolated process tracing a path through space and time—a process that could go from one point in spacetime to another without exceeding the speed of light, for example. This terminology is particularly prominent in discussions of special relativity, where events that are connectible by slower-than-light physical processes

---

[30] In one of his unpublished manuscripts on his causal theory dating from the 1680s, Leibniz wrote: "The *full cause* is a producer that is prior by nature to what is produced, that is, that which involves all the requisites that are sufficient (i.e. from which the remaining requisites follow)." (A VI 4, 564). A "producer" (*inferens*), on the other hand, is defined as "that which, when posited, the other thing is posited" (C 471). See Futch (2008) and Arthur (2016) for details.

[31] (A VI 4, 568); see Arthur (2016) for an exposition of Leibniz's causal theory of time.

are termed *causally connectible*. The significance of this (as we shall see in Chap. 6) is that if such a process can extend from an event *a* to an event *b*, then *a* is (absolutely) before *b*. Such a relation is invariant, and does not depend on the reference frame chose to represent the process in space and time coordinates. Alfred Robb and others showed how this feature can be taken as definitional, and from this assumption together with minimal symmetry requirements and the hypothesis of a maximum signal velocity, the Lorentz equations of special relativity may be derived. This delivers a causal theory of spacetime, which may thus be regarded as a modern fulfillment of Leibniz's causal theory of time.[32]

Now it is perhaps understandable that physicists should call events that are connectible by physical processes *causally connectible*, in that in order for something at one such point to cause something to happen at the second point by physical means there would have to be an influence, or some chain of influences, propagated from the first point to the second, and it would need to traverse some such path slower than light. But as Bunge has observed, such a pair of events need not be connected by a *causal* process at all. The operant principle here is the Principle of Retarded Action, according to which any physical process must take a certain finite quantity of time. There is no necessity for this process to be a causal one:

> The principle of retarded action is independent of the causal principle, and, whenever it is postulated, it entails a restriction on the possible genetic connectivity of the physical level of reality… [M]oreover, the mere statement of retarded action, far from being essentially committed to causality, is consistent with noncausal categories of determination. (Bunge 1979, 67–68)

For example, a non-causal process of becoming will still be consistent with the principle of retarded action. In fact, an inertial motion is a perfect example of such a non-causal motion, as Bunge observes (Bunge 1979, 110–111). An inertial motion is one that evolves deterministically from an initial state to a final one precisely in the absence of causation.[33] And, on the other hand (contrary to Leibniz's stipulation), instances of causation may be held to be instantaneous, as in Newtonian gravitation theory. So "causal connectibility" is actually a misnomer. The term is probably too well established to be changed now, but as Roberto Torretti has argued elsewhere (Torretti 2007a), such terminological infelicities may well breed conceptual confusion.[34]

A causal process that proceeds from one point to another by a chain of intervening changes constitutes a *causal chain*. Now if other things interfered with a causal process along the way, this chain of influences might not have its effect. So the chain must be an isolated process. As Bunge has observed, this is an under-appreciated fact about causal chains: they must be isolated processes. But a complete isolation, of course,

---

[32] See Winnie (1977) for details.

[33] As Steve Savitt has reminded me, however, there exists a literature on causation by lack of a particular action, in claims such as "Your not watering my plant while I was away caused it to die". I will not pursue that here.

[34] Torretti gives examples such as "rod contraction" and "particle". One might add "wave-particle duality" and the "uncertainty principle", as we shall see in Chap. 8.

can never occur in fact, because "actually an infinity of neglected factors—Galileo's *cause accidentarie* or *cagioni secondarie*—are constantly impinging on the main stream" (Bunge 1979, 130).[35] At best causal chains work "as rough approximations for short periods of time", and this can only work if we "assume a fictitious *isolation* of the process in question from the remaining processes" (Bunge 1979, 127).

So we have a paradox: "the fiction of an isolated 'causal chain' will work to the extent to which such an isolation takes pace" (Bunge 1979, 130). Yet this fiction is a methodological requirement for the application of causal ideas to reality: "The isolation of a system from its surroundings, of a thing or process from its context, of a quality from the complex of interdependent qualities to which it belongs—such 'abstractions', in short, are indispensable not only for the applicability of causal ideas but for any research, whether empirical or theoretical" (Bunge 1979, 129). Thus, on the one hand, "The picture of linear causal chains is ontologically defective because it singles out a more or less imaginary line of development in a whole concrete stream" (Bunge 1979, 132); and, on the other, "such a linear character of causation is not altogether fictitious; it does work in definite respects and in limited domains. Causal chains are, in short, a rough model of real becoming" (Bunge 1979, 147).

Bunge's remarks are very much of a piece with Leibniz's epistemology of science. For precisely this contrast between the inherent complexity of reality as it exists *in concret*o, and the necessity of making abstractions from that complexity in order to have any knowledge of it, is right at the heart of Leibniz's philosophy. This is evident in the following passage from his *New Essays on Human Understanding*:

> if we thought in earnest that the things we do not apperceive are not there in the soul or in the body, we should fail in philosophy as in politics, by neglecting το μικρόν, imperceptible changes; whereas an abstraction is not an error, provided we know that what we are ignoring is really there. This is the use made of abstractions by mathematicians when they speak of the perfect lines they ask us to consider, and of uniform motions and other regular effects, although matter (that is to say the mixture of the effects of the surrounding infinite) is always providing some exception. We proceed in this way so as to distinguish the various considerations from one another, and, as far as is possible, to reduce effects to their reasons, and foresee some of their consequences. For the more careful we are to neglect no consideration which we can subject to rules, the more closely does practice correspond to theory. (A VI 6, 57/Leibniz 1981, 57)

The causal connection between any two existing events depends on all these concrete conditions; and according to Leibniz, it is "the mixture of the effects of the surrounding infinite" that makes any such connection contingent. For the contingency of a given event in the world is a consequence of its being related to any preceding cause by an infinite chain of intermediate causes or intervening changes. Given his doctrine that a demonstration must be effected in a finite number of steps, the infinitude of these intervening conditions guarantees that it is impossible to demonstrate a neces-

---

[35] Here Bunge approvingly quotes J. D. Bernal: "chance variations or side reactions are always taking place. These never completely cancel each other out, and there remains an accumulation which sooner or later provides a trend in a different direction from that of the original system" (Bernal 1949, 31; Bunge 1979, 131).

sary connection between cause and effect.[36] Still, as Bunge has also stressed, under certain conditions we may abstract from all this intervening complexity to identify isolated causal chains and deliver models of real becoming. But we should not lose sight of the fact that these are only models.

Leibniz's contrast between the abstract nature of the causal chains of our mathematical models and the (perhaps) infinite complexity of the world from which we abstract them is largely lost in subsequent philosophical thought on physics. This is already the case with d'Alembert, who regards the physical world as constituted by the algebraic relationships of his analytic mechanics, a picture that is then further entrenched in the rational mechanics of Euler, Lagrange and Laplace. In Laplace's determinism, initial conditions are simply configurations of the system: if the position and velocity of every particle is known at some initial moment, together with the laws governing the whole universe, all subsequent configurations can be calculated with geometric necessity.[37] But the most important shortcoming of his account is his (and his contemporaries') failure to recognize that the deterministic laws of physics not only abstract away from an infinity of neglected accidental factors but also, in order to be applied, require detailed knowledge of initial conditions not contained in the laws themselves.[38] This has the greatest bearing on attempts to define the direction of time, as we shall now discuss.

4.4 Entropy and Time Direction

Our conception of the one-way flow of time, its irreversibility, is almost inevitably bound up with considerations of ageing and decay and their irreversibility. Aristotle noted how this has led to a tendency to ascribe to time a certain destructiveness, and reported some of the sayings current in his own time: "time wastes things away, and things grow old by time" (Aristotle, *Physics* Bk IV, ch. xii 221b). Destruction is held to be intrinsic to time, he explains, because "change in itself is a departure from an existing condition" (222b 20–27). Although things are also *generated* in time, that requires extrinsic causes, whereas it is this "kind of perishing without apparent provocation that we especially attribute to time" (222b 20–27; quoted from Barbour

---

[36] Thus if being analytically true means being capable of being demonstrated, then (given Leibniz's demand that a demonstration be completable in a finite number of steps), only necessary truths are for him analytic, even though all truths have a sufficient reason.

[37] Lee Smolin calls this "the standard methodology of physics", and ascribes it to Newton, calling it the "Newtonian paradigm". Here a theory is applied to a "subsystem of the universe, idealized as an isolated system". Its kinematics is described by "a state space, C, giving all the possible states the system may have at any moment of time"; the dynamics then consists in a law governing the evolution of this system from some point in C. "The state space and the law are timeless, while the law evolves the state in time" (Unger and Smolin 2015, 373). The novelty of Unger and Smolin's view is that they challenge the idea of timeless laws, holding that the laws evolve in a global time through the evolution of the universe. (See also Unger and Smolin 2015, 19–22, 373).

[38] For an eloquent exposition of a similar point of view about accidental factors, the qualitative infinity of nature, and the idealizations of physics, see Bohm's (1957), esp. pp. 132–160.

2001, 83). Still, he sagely observes, it should not be thought that time *causes* this continual departure from existing conditions; “rather, this change happens to come to be in time”. It is the spontaneous and constant passing away of things and their changes that constitutes the irreversibility we attribute to time.[39]

The association of time with ageing and decay, however, has reappeared with the modern association of time’s direction with entropy. The great majority of modern physicists and philosophers of science, observing with Max Born that “there is no intrinsic direction in the flow of time contained in the equations” (Born 1949, 30), have therefore sought some physical basis for distinguishing one of the two putative directions of time. And the obvious candidate for such a criterion seems provided by the fact that, according to Clausius’s formulation of the Second Law of Thermodynamics, *entropy never decreases in time*: irreversible processes are characterized by an increase in entropy, while in reversible processes it remains constant.[40]

There is no doubt about the relevance of entropy to a great many issues of concern to ourselves and our knowledge of the universe—everything from the exhausting of usable fuel resources on Earth to Hawking radiation from black holes. But, I maintain, it is a mistake to try to define the direction of time in terms of the tendency of entropy to increase. I do not believe that there are in fact *two* directions of time. All concrete processes develop from earlier to later, with their earlier states “perishing” as they proceed to the later ones, and it is in this earlier-to-later order of changes of state that the direction of time consists.[41]

As a first step towards making a case for this, let us consider a scenario proposed by Karl Popper, which he offered as a “trivial counterexample” to the claim that “all irreversible mechanical processes involve an increase in entropy” (Popper 1956, 538). It appears to have been prompted by Max Born’s discussion of Maxwell’s theory. For as Born notes, the solution of Maxwell’s equations given by so-called retarded potentials corresponding to an electromagnetic wave spreading from a point source is matched by a time-reversed solution given by the advanced potentials which represents a wave contracting towards the source. But the latter solution is only applied in very special circumstances, such as a spherical wave being reflected by a concentric spherical mirror. So in general the equations must be “supplemented by the rule that only retarded solutions are allowed” (Born 1949, 26). According to Born, this “principle of antecedence” has to be put in by hand, as it were. In a similar vein, Popper suggests the example of a film being made of a large stone being dropped onto the still surface of a pond. “The reversed film will show contracting circular

---

[39] That “the irreversibility of time is the foundation of the asymmetry between past and future”, and not derived from entropic considerations, is also contended by Roberto Mangabeira Unger (Unger and Smolin 2015, 235).

[40] “One can express fundamental laws of the Universe that correspond to the two main laws of thermodynamics in the following simple form: (1) The energy of the Universe is constant. (2) The entropy of the Universe tends to a maximum.” (Clausius 1867, 44, as quoted in English by Uffink 2003, 129; see Torretti 2007b, 738).

[41] Here I am in complete accord with Mauro Dorato: “physics cannot provide *empirical* evidence for the reality of absolute becoming because it presupposes it, at least to the extent that it presupposes an ontology of events” (2006, 569).

waves of increasing amplitude", until finally "immediately behind the highest wave crest, a circular region of undisturbed water will close in towards the centre". As he observes, such a reversed sequence "cannot be regarded as a classical process", so it must be conceded that "irreversible classical processes exist" (Popper 1956, 538). This example, together with the example of choosing the retarded potentials in Maxwell's theory, indicate that "although the arrow of time is not implied by the fundamental equations, it nevertheless characterizes most solutions". And the reason it is implied in the solutions, I argue, is that the equations must be applied to concrete situations in which processes proceed *from initial* conditions, a circumstance that necessarily presupposes that processes are oriented in time from earlier to later.[42]

We can illustrate the same point with a simpler example. Imagine a break in a game of billiards or pool. The cue strikes the cue ball and propels it down the table, where it hits the ten balls arranged in a triangular rack. Neglecting friction, the balls all rebound according to the time-symmetric laws of classical mechanics. Assuming all the collisions are elastic, the time-reverse of this collision is also covered by the same laws. If we were to film the process, and play the film backwards, there'd be no outright incongruity. It would, however, be extremely hard to replicate the reverse break in real time, since it would require an amazing conspiracy of precisely the right initial velocities and directions of all the various balls in order to end up with the rack of balls stationary and the cue ball shooting back down the table and coming to a rest against the reversing cue.[43] It is the laws together with the initial conditions that make one scenario so much less likely than the other. Similarly, if gas molecules in a container are rebounding from one another elastically, a configuration where all the molecules are initially on one side of the container would appear to be very much less likely than one in which they are all more or less randomly distributed.

It was on such considerations about probability of configurations that Ludwig Boltzmann developed statistical mechanics. For him, however, configurations are conceived entirely in terms of velocities and positions of constituent particles, in keeping with his programme of reducing phenomenological thermodynamics to statistical mechanics. In particular, he sought to derive the Second Law of Thermodynamics, introduced by Clausius, according to which in a heat exchange entropy never decreases, but can only increase (in irreversible processes) or stay the same (in reversible ones). The latter fact, Boltzmann then argued, provides an explanation of "time's arrow": *the direction of time is to be identified with the direction of increas-*

---

[42] Thus my point of view is directly contrary to that adopted by Huw Price in his provocative (1996). Price insists on a "block universe" view (version 1) which denies the flow of time, so that there is no objective difference between "the two directions of time". Thus, in his analysis of Popper's argument, and throughout his book, he assumes that "in changing the perspective [from one sense of time to the other] we do not change anything objective" (56).

[43] Here it might be objected that the configuration of all the balls in the first scenario is equally improbable, since it takes an exquisite precision in the velocity and direction of the cue ball to result in precisely that particular break. Carlo Rovelli has made this point, arguing that all such scenarios are equally particular. But in a realistic depiction of the pool break, we cannot ignore the fact that the system of moving balls not isolated from its surroundings: friction cannot be neglected, and some of the balls may have dropped into pockets. This lack of causal isolation is what makes the initial conditions in the reverse case so stupendously improbable.

ing entropy itself. The macrostate corresponding to equilibrium is overwhelmingly more probable than other states, so that if the universe had begun in a highly ordered and therefore less probable configuration, it would proceed inexorably towards a state of thermal equilibrium, a state of maximal entropy or "heat death". On this view, the fact that we experience time as going only one way is a result of the fact that our experience itself is an irreversible process aligned with the other irreversible processes in our domain of experience.

Boltzmann's work in statistical mechanics is justly celebrated. It cannot be regarded as a successful reduction of thermodynamics to statistical mechanics, however, and Boltzmann's claim to have derived the time-asymmetric Second Law from time-symmetric mechanical laws, and his subsequent attempts to clarify his position, have been beset by insuperable objections from the outset. It is therefore surprising that so many physicists and philosophers take for granted a Boltzmannian foundation for the Second Law. Even more surprising is the widespread acceptance of his application of statistical dynamical analysis to the universe as a whole, and his subsequent declarations that it must have begun in a low entropy state, and that the direction of time must be defined as the direction of increasing entropy. This is not the place to give a thorough critique of Boltzmann's reductionist programme, or to point out the many difficulties that afflict it. That has in any case already been very capably done elsewhere.[44] After giving a rough sketch of the main features of Boltzmann's theory, and some of the objections that have been made against it, I shall attempt my own assessment of the reasons for its failure. Essentially, it is misconceived insofar as it tries to derive irreversibility from the *laws* of physics alone; the conditions Boltzmann assumes for the application of his analysis are not met in its application to the universe as a whole; and the attempt to *define* the direction of time in terms of increasing entropy is incoherent.

In broad strokes, the phenomenological Second Law of Thermodynamics developed out of the concern to make heat engines as efficient as possible. It was well known that when heat is converted to mechanical energy as in, for instance, a steam engine, some energy is always lost to the environment through friction and other heat losses; there must be a positive temperature difference in the heat source in order for energy to be converted into mechanical work.[45] This was quantified by Rudolf Clausius in terms of a new concept entropy S, by defining the change in entropy ΔS of a heat exchange ΔQ at an absolute temperature T as $\Delta S = \Delta Q/T$. The idea is that entropy is conserved for a reversible process, but inevitably lost in an irreversible process. This gives the phenomenological Second Law: $\Delta S = \Delta Q/T \geq 0$.

Boltzmann proposed that the irreversibility sanctioned by this phenomenological law is not absolute, but statistical in origin. He imagined an ideal gas, that is, one

---

[44] See for example Earman (2006), Uffink (2007), Butterfield and Earman (2007) and Torretti (2007b).

[45] "It is impossible for a self-acting machine, unaided by any external agency, to convey heat from one body to another at a higher temperature"—Rudolf Clausius, quoted from Torretti (2007b, 740). Torretti also notes the alternative formulation by Flanders and Swann: "Heat won't pass from a cooler to a hotter. You can try it if you like, but you far better notter!" —quoted from "At The Drop Of Another Hat" (1964).

constituted solely by identical molecules that rebound from one another perfectly elastically without interacting, and whose states at any time are completely characterized by their velocity and position alone. He was able to show that the entropy S of the gas would be given by the expression $S = -k \log W$, where W is the probability of some macroscopic state of the gas for some possible distribution of microstates, 'log x' denotes the natural logarithm of x, and k is now called Boltzmann's constant. A microstate is a system of N particles, each having an exact value of canonical variables such as position and momentum (in each of three directions) at each instant, giving a configuration of $6N$ coordinates; the set of all such configurations constitutes the system's *phase space*. A *macrostate* is a set of microstates giving rise to the same values of such observable macro-properties as volume and pressure (in the lingo, this is called the *coarse-graining* of the gas). The probability of a macrostate is then expressed as a measure function on the phase space: the more ways in which a given macrostate can encompass different microstates, the more probable it is. Entropy increases just because a highly ordered macrostate—such as one where the molecules of the gas are all on one side of a container or all moving in the same direction—is far less probable than one where they are more or less equally distributed throughout the container and moving in randomly distributed directions. Processes involving a vast number of microstates will develop from more orderly (and therefore less probable) initial configurations to less orderly (and therefore more probable) ones.

Boltzmann's view ran into problems, not the least of which is Loschmidt's paradox, also called the reversibility paradox.[46] This is the objection that it should not be possible to deduce an irreversible macroscopic process from laws of dynamics that are time-symmetric. Loschmidt pointed out that if you have a process consisting in a system increasing in entropy as it goes from state s_1 to state s_2, then, because of the time-symmetry of the laws, a process for that system going from s_2 to s_1 with all the velocities reversed, would be equally probable. Moreover, the initial conditions for the second process, if picked randomly from the phase space of all possible states for that system, would be just as probable as the initial conditions for the first process. This objection is still widely accepted among contemporary physicists and philosophers.[47]

---

[46] The objection was raised by Loschmidt in his (1876), and supported by the arguments of Burbury in 1894, and Zermelo in 1896. See Torretti (2007b) for discussion and references to original papers and recent discussion by philosophers of physics. Torretti summarizes the latter as "sufficient to dismiss the popular understanding of the second law of thermodynamics as a law of cosmic evolution, to disqualify thermodynamic entropy as the physical source of universal time order, and to remove the need for deriving Time's Arrow—*per impossibile*—from the mechanical or statistico-mechanical principles of thermal physics" (2007b, 739).

[47] For example, this is the explanation Sean Carroll gave to journalist Dan Falk, who was covering a conference on Time in Cosmology at the Perimeter Institute in Waterloo, Ontario: "What Boltzmann truly explained is why the entropy of the universe will be larger tomorrow than it is today. But if that was all you knew, you'd also say that the entropy of the universe was probably larger yesterday than today—because all the underlying dynamics are completely symmetric with respect to time." (Falk 2016).

One thing we may take this to indicate is that the initial conditions for an irreversible process cannot be determined by looking at one system in isolation, with its phase space of reversible microstates. That is, each irreversible process (such as a billiard ball dropping into its pocket) will involve a transfer of energy (and possibly matter) across a divide between two systems that are precisely not isolated from one another. The initial conditions for this process may be determined by the ball's position and velocity, but those of the reverse process, which involve the conspiring of countless micro-motions in the pocket, returning sound waves, and so forth, are determined by the highly contingent situations and properties of those various particular subsystems as well as an input of energy that would violate the condition of causal isolation.[48] Boltzmann's ideal gas model assumes that there is one isolated system containing subsystems that are reversible processes, and that initial conditions are simply configurations of this system. A realistic model would recognize that, although motions might be time reversible, each subsystem is an irreversible process involving a transition from less probable initial conditions to more probable ones. These conditions are entirely contingent: think again of the arrangement and input of energy (violating the condition of causal isolation) necessary for the billiard balls all to jump up out of their pockets with the needed velocity, spin and direction of motion. Once those initial conditions are in place, the subsequent motions are reversible, but the reverse process, including these conditions, is almost infinitely improbable: the original process is therefore irreversible. Consequently, reversing a sequence of such irreversible processes would produce a sequence of processes each involving a transition from more probable initial conditions to infinitely less probable ones. The appearance of absurdity when we play a movie of such processes backwards is thus an objective feature of reality, and not a function of the fact that our memories are aligned with one preferred direction of time. This, I suggest, is the lesson to be drawn from Popper's example of the reversed film of a pebble thrown into a pond.

Further objections can be made to Boltzmann's assumption that his ideal gas model would be applicable to the universe as a whole. The ideal gas of his model consists of molecules undergoing elastic collisions and confined in a finite space by rigid walls. The molecules are assumed to have the same probability of moving in any given direction, and to have kinetic energies that are initially uniformly distributed and remain so. But, of course, there are no container walls in space, the galaxies are not bouncing off one another, nor are there any regions of space containing galaxies whose energies are uniformly distributed.[49]

---

[48] This is related to a reply S. H. Burbury made in a letter to *Nature* in 1894 in defence of the statistical increase in entropy (Boltzmann's H-function) against the reversibility objection: "I think the answer to this would be that any actual material system receives disturbances from without, the effect of which, coming at haphazard, is to produce that very distribution of coordinates which is required to make H diminish" (Burbury 1894, 78).

[49] Perhaps (as suggested to me by David Wright in private correspondence) clusters of galaxies can be considered as confined in a gravitational potential well, thus mimicking the effect of confinement by a wall. But for detailed discussion and criticism of these and other assumptions made by Boltzmann, see Earman (2006), Uffink (2007), Butterfield and Earman (2007) and Torretti (2007b).

Boltzmann's own response to Loschmidt's paradox is worth pursuing a little further, however, despite its lack of realism, because of its influence on subsequent debate about the direction of time. He reasoned that if the universe itself is an isolated system, then although as a whole it would most likely be in an equilibrium state, there would be fluctuations on either side of equilibrium. He then assumed that if we were in a region of the universe where the entropy had fluctuated to a state of higher order, then the entropy in this region would be increasing as the systems in the region tended back towards equilibrium.[50] The fact that we experience time going forwards is due to the fact that we are in such a region of the universe. If we were in a region where entropy were decreasing, time would go "backwards", although we would not be aware of this, since we would share the same direction of time.[51] Following this logic, the philosopher of physics J. J. C. Smart argued that, accepting Loschmidt's objections, it would follow that "any evolutions of this universe would eventually be followed by precisely the reverse devolutions" (Smart 1968, 263). That is, if in the future the same sequence of cells in phase space were to occur in reverse order, then "people like us would be performing exactly the same actions, and having exactly the same stimuli, except backwards in time" (263).

Of course, all this is extremely contrary to fact. The universe as we know it is not a thermodynamic system in a state of near equilibrium: in fact, as Larry Sklar observes, it is "in a highly non-equilibrium state with parallel entropic increase into the future everywhere" (Sklar 2015, §5). Moreover, as John Earman has objected, it "is homogeneous only at large scales—above 10 megaparsecs—and at the smaller scales at which we typically apply thermodynamic reasoning, the inhomogeneities can be expected to be associated with entropy gradients" (Earman 2006, 413). Processes such as stars shining, the Sun heating the Earth, the formation of galaxies, the generation of living organisms, and so on, are all irreversible processes—and therefore not analogous to the reversible microprocesses assumed in Boltzmann's ideal gas model. His reasoning, moreover, is based on the nineteenth century conception of a universe that is indefinitely old and finite in size, containing a homogeneous distribution of galaxies. But as we now know, because of the expansion of space from the hypothesized Big Bang, the universe is finite in age, thus putting some of Boltzmann's constitutive assumptions into doubt.

The very fact of the expansion of space, however, has been interpreted as providing the basis for the asymmetry in the direction of time that Boltzmann was unable to provide with his classical model. On this view, the expansion of space against the

---

[50] "Then in the universe, which is in thermal equilibrium throughout and therefore dead, there will occur here and there relatively small regions of the same size as our galaxy (we call them single worlds) which, during the relatively short time of eons, fluctuate noticeably from thermal equilibrium, and indeed the state probability in such cases will be equally likely to increase or decrease. For the universe, the two directions of time are indistinguishable, just as in space there is no up and down." (Boltzmann 1964, 446–47).

[51] "However, just as at a particular place on the earth's surface we call 'down' the direction toward the center of the earth, so will a living being in a particular time interval of such a single world distinguish the direction of time toward the less probable state from the opposite direction (the former toward the past, the latter toward the future)." (Boltzmann 1964, 447).

action of gravity produces a constant source of available potential energy: matter that has been forced apart by the expansion will thereby have gained potential energy that can be reconverted into motion (and useful energy) as it falls back together. Such an energy exchange can produce increasing order in the target systems (e.g. life on Earth), even though it is itself a far-from-equilibrium process. Therefore if matter and energy were in the early moments of the universe distributed fairly uniformly, then (given the expansion of space against gravity) this would be a state of lower entropy than exists now. So on this interpretation, the second law of thermodynamics has its ultimate basis in the expansion of the universe. Thus, as Thomas Gold was the first to propose (Gold 1962, 1966), if the direction of time is given by the direction of increase of entropy, then time's direction is an effect of the expansion of the universe.

But the appeal to cosmology to save Boltzmann's program is much more problematic than might be supposed from its widespread endorsement. It depends crucially on the contentious assumption, usually called the "Past Hypothesis", that the universe began in a state of extremely low entropy. On close analysis, however, the Boltzmann entropy of the initial state of the universe seems "to be an ill-defined or severely hobbled concept" (Earman 2006). Worse, the proposal that it should be a state of extremely low entropy appears to be at odds with the evidence provided by the *cosmic microwave background* radiation (CMB)—itself a vital piece of evidence for the Big Bang. The CMB is radiation due to the decoupling of matter and radiation in the extremely hot early universe, estimated to have occurred at about 379,000 years after the Big Bang when the universe would have had a temperature of about 3000 K. Because of the expansion of space the wavelength of this light would have been stretched out, lowering its frequency and energy so that it would now correspond to a temperature of about 2.725 K. Observations show radiation corresponding to exactly this temperature coming from all directions in space. But here is the kicker: these very observations show this radiation to conform precisely to "Black Body radiation", characteristic of a state of thermal equilibrium. This, however, is a state of maximum entropy! As Roger Penrose writes, "Thermal equilibrium represents the macroscopic state that one envisages a system finally settling into, sometimes referred to as the *heat death of the universe*.... The CMB evidence seems to be telling us that the macroscopic state of the Big Bang has an enormous entropy, even equal to the maximum among all possibilities" (Penrose 2016, 252).[52]

---

[52] For a thorough critique of the "Past Hypothesis", see Earman (2006). As he and Torretti (2007b) have objected, the hypothesis of an initial low entropy of the universe is at odds with the stringent condition of homogeneity that must be satisfied by the FLRW models of spacetime (see Chap. 7) on which speculations about the early universe are based. It is "*only* if the distribution of energy on each hypersurface of simultaneity is absolutely uniform" that the Einstein field equations admit FLRW models as solutions (Torretti 2007b, 751). There are also problems that I will not go into here associated with the fact that the expansion of the universe is accelerating: see Earman (2006, 413). Penrose, however, argues in his recent book that "when gravity is brought into the picture, the CMB must actually have been very far from a maximum-entropy state" (Penrose 2016, 255), even if it issued from a state of thermal equilibrium. This is because "there can be an enormous entropy gain once we allow significant deviation from spatial uniformity, the greatest gain arising from those irregularities leading to black holes" (2016, 255).

The explanation for the direction of time is terms of expansion of the universe is, however, vulnerable to other damning objections. What if the universe began to contract? Surely, on this view, the direction of time would be inverted? Granted, such a contraction now appears to be ruled out by recent observations of the increased rate of expansion, but we are dealing here with matters of principle. So the question is, would the direction of time be reversed if the universe began to contract? Would all processes start occurring in the reverse direction of time? Stephen Hawking believed they would when he first proposed his "no-boundary condition" theory—a cosmological theory combining the idea that there are no boundary conditions for the universe as a whole, with the weak anthropic principle.[53] He recalls that he initially believed that this theory "did indeed imply that disorder would decrease in the contracting phase" (Hawking 1988, 150). He had, in his own words, assumed that "the contracting phase would be like the time reverse of the expanding phase. People in the contracting phase would live their lives backward: they would die before they were born and get younger as the universe contracted" (150). But is such a scenario even coherent? What would happen at the point of reversal, when the universe finally stops expanding and begins to contract?

One can already note a potential paralogism in the very supposition of a "point of reversal". On the one hand, the entropy would presumably reach a maximum when the last work-convertible energy in some extremely localized region were used up; but the change from expansion to contraction would be a global change, affecting all of space at a given time, assuming there is such a thing as a global space at a given time. Suppose, that is, that we have a spacetime whose time orientation is defined by the direction of increasing entropy of the universe as a whole: i.e. a cosmic time function that runs with the expansion, but reverses direction at the point of change from expansion to contraction. Suppose one of the last processes with time running forwards to be a particle moving inertially with velocity v. Now what happens when the direction of time is reversed, that is, when $t \to -t$? If nothing else changes, $v \to -v$. Indeed, all particle velocities will be instantaneously reversed. This seems crazy.[54] One could perhaps prevent discernible causeless changes of this kind by requiring such processes to suffer motion reversals simultaneously. But even putting aside difficulties about simultaneity of distant events that we will discuss in later chapters, this scenario seems to be one completely lacking an independent motivation in the physics.[55]

I think there are two lessons to be learned here. One is that the direction of time is not *defined* by the direction of increasing entropy. The second is that the directionality

---

[53] This is the methodological principle that the universe must have the properties necessary for the emergence of living, conscious observers, given the obvious fact of their existence.

[54] Cf. Torretti, who, on having just seen two swallows flying past his window, asks "Would I see the birds' heads trail their bodies if I lived in a 'single world' in which entropy was decreasing?" (Torretti 1999, 213).

[55] Another objection raised by Penrose concerns the formation of black holes with their concomitant massive increase in entropy. If it is assumed that increasing entropy defines the direction of time, then reversing the direction of time would require the reversing of black hole formation, contrary to accepted physics (Penrose 2016, 254).

in time of individual processes does not depend on the global direction of time being determined by the overall increase in entropy.

The first lesson can be clarified by an analogy. Imagine swallows gathering on telephone wires to be an indication of the beginning of summer. If the swallows failed to gather one year—perhaps because of a change of migration route or a global climate change, or whatever—we would not take that as showing that summer had not begun. The gathering of swallows is merely a criterion or marker of a seasonal change that we take for granted, not a defining factor. (Cf. Aristotle's old saw: "One swallow does not a summer make"!) Similarly, increasing entropy is not constitutive of time's direction. It characterizes an isolated system's going from one state to another of greater disorder. If we discover this to be the case, then we can be pretty sure that the more disordered state is the later of the two. But if this is not so, either because the processes leading from one state to the other are reversible, or because the entropy decreases (a statistical possibility according to Boltzmann's theory), we cannot thereby conclude that time's direction has changed. In fact, even in order to determine whether the entropy increases or decreases in a given process, we have to assume a time direction for that process. The conclusion is inescapable: the direction of time cannot in principle be *defined* as the direction in which entropy increases.

The same considerations imply, secondly, that the directionality in time of individual processes does not depend on a global direction of time. An irreversible process could still be locally irreversible because of the overwhelming improbability of the reverse process proceeding from its final to its initial conditions, even if there were no time function defined in terms of the expansion of the universe. In principle, we could live in a spacetime that is not time-orientable, i.e. for which no global time function is definable. But the spacetime could still contain irreversible processes, and these would still have initial and final states. They would therefore still be locally time-oriented, even if there were no way to stitch them all together so that a consistent global time direction could result.[56] Whether such a universe would be consistent with the reality of becoming, however, is another matter, to which we shall return in Chap. 7.

4.5 Time Symmetry and Time Reversal

But there is a deeper point to be made. This concerns the very meaning of time reversal. It is usually assumed that replacing t by $-t$ in the physical laws or equations governing types of processes is equivalent to reversing the direction of time—that is, to reversing the order of succession of the actual events or states making up the processes. But these are not the same thing. Take, for example, the usual example that is used to display the one-wayness of time, the absurd sequences displayed when a movie is run backwards. Here it is of course only the sequence of frames of the movie

---

[56] See Earman's discussion in his (1974). As he notes, if a time direction is supposed for "sufficiently small regions of spacetime" rather than globally, then "this presupposition always holds" (33).

that is reversed with respect to the standard time direction. We who are watching it, the projector, and the frames appearing to us are all processes running "forwards" in time. In fact, since the direction of time is just the direction of process, if time is reversed then all processes are reversed, and this would be completely indistinguishable from the normal state of affairs. So "time reversal" is really a misnomer. It would be more accurate to call it (as is sometimes done) "*motion* reversal". But even this does not quite capture it. The point is that the events that make up a particular process—the process-tokens—occur only in the order that they do, with the later events coming out of the earlier ones. A reversible process is one in which event tokens of the same *type* of process may occur in the reverse order with respect to the standard direction of time exhibited by the processes surrounding them. An irreversible process is one in which event tokens of the same *type* of process cannot occur in the reverse order with respect to that standard direction of time.

The laws of physics, that is, concern *types of processes*. If the equation governing a type of process is time-reversal invariant, this does not mean that the same individual process can occur going either forward or backward in time. It means that a process running through a sequence of types of states $[s_0, s_1, s_2, \ldots, s_{i,} \ldots, s_{n-1}, s_n]$, is just as likely to occur in reverse order in time, $[s\dagger_n, s\dagger_{n-1}, \ldots, s\dagger_{i,} \ldots, s\dagger_2, s\dagger_1, s\dagger_0]$, where the $s\dagger_i$ are the time-inverses of the original states s_i. Thus if s_i. is the instantaneous state of a particle with a clockwise spin and an instantaneous velocity v directed towards the left, then $s\dagger_i$ is the state of a particle with the same intrinsic properties but with a velocity $-v$ towards the right and an anticlockwise spin.[57] For them to occur in reverse order is not for *time* to be reversed, but for their sequencing to be reversed with respect to the standard order in time exhibited by all other processes.[58] In fact, it turns out to be a fundamental fact that if a given process is subjected to the three transformations of time reversal, inverting direction in space (parity reversal), and charge conjugation (turning a negative charge into a positive one, and vice versa), the result is the type of process with which you began. This is the CPT theorem, embodying a fundamental symmetry of physical theory.

This symmetry is the basis for a radical theory of antiparticles proposed by John Archibald Wheeler in the early 1940s. Since a positron has opposite charge and spin to an electron, the time inverse of the process of an electron moving through spacetime would be indistinguishable from the motion of its antiparticle, the positron. This led him to conjecture that "a positron could be interpreted as an electron moving backward in time" (Wheeler 1998, 117). "I know why all electrons and all positrons have the same mass and charge!", he excitedly told his graduate student Richard Feynman: "They are the same particle!" (117). Wheelers' conjecture does not pan out, since in the observable universe antiparticles are massively outnumbered by particles. Nevertheless, the basic idea was incorporated by Feynman into his famous diagrams

---

[57] In quantum theory, there is the added condition that the time inverse of a state must be the complex conjugate of the original state.

[58] Cf. Earman (1974, 37): "*We live in only one model, and any given model can be as radically asymmetric with respect to past and future as you like while at the same time all of the relevant laws are time reversal invariant*" (italics in the original).

that are used for computations in quantum electrodynamics, where the annihilation of positron on colliding with an electron to produce a photon is interpreted as an electron being knocked backward in time by its collision with a photon. Such a reinterpretation is, in my opinion, untenable. Granted, it does not involve the same fallacy as that committed by H. G. Wells in having his Time Traveller journey back and forth at various speeds along the time axis as though it were a line in space. But it still treats a path through time as though it were a path through space that can be traversed in either direction, as suggested by the representation of a four-dimensional worldline in a spatial diagram. The times at which a positron is at different places in space are in relation to the times of other events occurring simultaneously with its being in those places: it does not have its own private time.[59] So the ordering of its states as concrete tokens cannot be reversed relative to the order in which everything else is occurring.

Similar misconstruals of time reversal abound in the literature. One reads that because the laws governing microprocesses are time symmetric, there is no direction of time at this level, and it emerges only in macroscopic processes.[60] But there is no way for microprocesses to compose into macroprocesses unless they are all evolving in time with their states in synchrony, at least locally. Likewise, it is supposed to be an open problem how the descent of matter into a black hole could be irreversible when all the laws of physics are reversible.[61] But this is to confuse the order of succession of concrete events or states with the order in which types of process may be oriented with respect to that order of succession. Not to put too fine a point on it, the idea that time reversal is a reversal of the order in which actual events succeed one another is on a level with the idea of Superman reversing time's direction by spinning the Earth in the opposite direction.

In sum, time-reversal does not involve the reversal of the order in which states or events become relative to the order in which other states or events things become. And conversely, the invariance of equations under time reversal, time reversal symmetry, does not entail the symmetry of the temporal order of concrete events or states. The

---

[59] It is often held that in relativity theory *proper time* constitutes such a private time for an observer along its worldline—e.g. "Each observer has his own *proper time*" (Morris 1985, 157). But as we shall see in Chaps. 5 and 6, although proper time is specific to a given trajectory in spacetime, it is a constitutive assumption of Minkowski spacetime that processes cannot reverse their temporal orientation.

[60] This has been the accepted wisdom for decades. Thus Richard Morris in 1985: "The seemingly paradoxical notion that time-reversed motion might be possible is a consequence of the fact that there is no arrow of time on the subatomic level" (Morris 1985, 126). At least Rovelli, who argues that "In a microscopic description, there can be no sense in which the past is different from the future" (Rovelli 2018, 33), allows that the "thermal time" that he takes to describe macroscopic phenomena also possesses no direction, "and lacks what we mean when we speak of its flow" (142). According to him, "The difference between past and future … issues only from the fact that, in the past, the world found itself subject to a state that, with our blurred take on things, appears particular to us" (194).

[61] Thus George Musser writes in a recent edition of *Nature* concerning descent into a black hole: "The descent is irreversible. That is a problem because all known laws of fundamental physics, including those of quantum mechanics as generally understood, are reversible." (Musser 2018, S4).

order in which processes occur and things become is intrinsically asymmetric. This would still be so even if all processes were symmetric under time-reversal.

Perhaps an analogy with language will help. In Arabic one reads from right to left, in European languages we read from left to right; no matter, choose one, say English. (This corresponds to what is conventional in the direction of time: by analogy with our choice of writing direction, we choose to represent the direction of flow as $+t$; it could equally well be chosen as $-t$.) Now the analogy with a type of process that can occur in reverse order is a *palindrome*, a word that reads the same forwards as backwards, like ‘level’ or ‘rotor’. Note that we still read the word from left to right; it is a palindrome if and only if it is indistinguishable from its “write-reverse”, that is, the same word rewritten so that its last letter is first, its second-last letter second, and so on. The important point is that we have to read sentences in a given direction in order to make sense of them. The fact that individual words could be read backwards does not alter this.

Those who wish to claim that the direction of time is a contingent feature of our temporal experience (grounded in the existence of irreversible processes constituting that experience) are committed to the view that in a world without such processes there is no direction of time. Let us suppose, then, counterfactually, that the universe consisted exclusively of reversible processes. The analogy here would be a language in which every word is a palindrome—certainly a possibility, it would not affect semantics or word order. Obviously we can create such a language trivially by simply adding on to the end of each word of, say, English, its write-inverse: Call this language PalinnilaP. HereereH issi anna exampleelpmaxe offo aa sentenceecnetnes.

Now one thing that is immediately obvious from this example is that a language in which every word is a palindrome—the analogue of a universe in which every process is reversible in time—does not itself display a reading symmetry (analogue: temporal symmetry): we still read it from left to right in order to understand it properly, just as temporally symmetric processes are still oriented in the same temporal direction as all other processes. And this order should not be thought of as purely conventional: once we have decided that reading and writing are to be done from left to right (this is the conventional element), we must read it from left to right in order to make sense of it. Similarly, once we have decided that processes are oriented towards positive t, applying the operation of time-reversal to a given type of process does not make that process go backwards in time, but reverses the order of the states of the type of process with respect to the usual time-orientation.

According to this analogy, each word of the language PalinnilaP would be write-reversible, i.e. palindromic, without the language as a whole (or even texts of it) being so; and the asymmetry derives from the fact that, in order to understand parts of the language, we must read all the words in only one direction. So, by analogy, each different type of process could be time-reversible, without this entailing that the universe as a whole is time-symmetric; and the asymmetry derives from the fact that, in order for the processes to occur at all (whether they are time-reversible or not), they must occur in the same direction, which, according to our conventions, is past-to-future.

Huw Price has challenged the cogency of such an argument. He believes that since it is a matter of convention whether we assume one or the other direction as *the* direction of time, this concerns the orientability of spacetime—a necessary condition for the direction of time, but not a sufficient one. "Orientability ensures that if we decide by convention that one of two asteroids is travelling from Mars, towards the Earth, then we can extend the convention to the rest of spacetime in a consistent way" (Price 2011, 215). To this it must be responded: we do *not* decide by convention which way an asteroid is travelling in space or in time. As a concrete process, it must proceed from some initial position to subsequent ones. There is nothing conventional about this.

To reiterate the main contention of this chapter: becoming takes place at the level of individual local processes that run from initial to final states, a circumstance that is presupposed in determining whether or not a type of process is time-reversible. A process is something that occurs from an initial state or event to some final state or event: the order is the (local) order of becoming of the states, and this is intrinsically asymmetric.

4.6 Summary

- In this chapter I have critiqued that the idea that there are several arrows of time, arguing that the quest to define time's arrow by a construction from laws is quixotic. Becoming takes place at the level of individual local processes that run from initial to final states, not at the level of laws; and this real, local passage is the basis of time's direction. Given this identification of the direction of time with the intrinsic directionality of process, I argued that the prevalent idea that there are two possible directions of time is mistaken.
- To make this case I began with a consideration of how becoming was understood in classical physics, comparing the views of Newton with Leibniz. I argued that despite the unsoundness of Newton's conception of time generated by a moving now, his notion of an equable flow still has empirical content. It corresponds to the time beaten out by a body moving inertially through space. I also argued that although both thinkers conceived of the direction of time in terms of time flow, Leibniz's conception is not susceptible to the same difficulties as afflict Newton's. This led into a consideration of the causal theory of time's direction, and objections that have been made to it.
- I argued first that causal determination is only one type of determination, and that that the directionality of causal (and other deterministic and indeterministic) processes results from the temporal directionality of process, rather than time's direction resulting from cause. A process is something that occurs from an initial state or event to some final state or event: the order is the (local) order of becoming of the states, and this is asymmetric. I also argued that concentration on causally deterministic laws crucially leaves out of account the fact that application of such laws to the world always involves appeal to initial conditions as well as laws.

- The consideration of initial and final conditions led naturally into a consideration of Boltzmann's theory that the direction of time be defined as the direction of increasing entropy. Objections such as Loschmidt's paradox were considered, and I argued that these share with Boltzmann's ideal gas model the unrealistic assumption that there is one isolated system containing subsystems that are reversible processes, and that initial conditions are simply configurations of this system. I argued that the irreversibility resides in what constitute the initial conditions of individual processes making up such a complex system, not in the laws governing the subsystems; and that, more importantly, time's direction cannot coherently be *defined* in terms of either increase in entropy or the expansion of the universe.
- Finally the arguments against passage from the time-reversal symmetry of fundamental physical laws were examined. It was argued that even in a universe in which all the fundamental laws are symmetric under time reversal, there would still be an intrinsic asymmetry of process. Elementary processes are processes, so they go from initial to final states, and are therefore intrinsically directed. To say that time reversibility means they can "go either way in time" confuses (a) the idea that the states of a given type of process can be ordered either way with respect to the standard direction of time, with (b) the idea that a concrete process can run either from past to future or from future to past with respect to that ordering, which it can't. It is also to mistake the time symmetry of a type of process for the symmetry of time itself: time reversal symmetry does not entail the symmetry of the temporal order of concrete events or states. If elementary processes were not oriented in time in the same way, they could not compose into a macroscopic process that is so oriented: the composition would not work. Thus time direction is not an emergent macroscopic property requiring a foundation in the laws of physics.

References

Arthur, Richard T.W. 1985. Leibniz's Theory of Time. In *The Natural Philosophy of Leibniz*, ed. Kathleen Okruhlik and James Robert Brown, 263–313. Dordrecht and Boston: D. Reidel.

Arthur, Richard T.W. 1995. Newton's Fluxions and Equably Flowing Time. *Studies in History and Philosophy of Science* 28 (2): 323–351.

Arthur, Richard T.W. 2016. Leibniz's Causal Theory of Time Revisited. *Leibniz Review* 26: 151–178.

Barbour, Julian B. 2001. *The Discovery of Dynamics*. [An edited reissue of *Absolute or Relative Motion? Vol. 1: The Discovery of Dynamics* (Cambridge: Cambridge University Press, 1989)]. Oxford: Oxford University Press.

Barbour, Julian B. 2008. The Nature of Time. http://www.platonia.com/nature_of_time_essay.pdf.

Barrow, Isaac. 1670. *Lectiones Geometricae*. London; reproduced in (Barrow 1860).

Barrow, Isaac. 1860. *The Mathematical Works of Isaac Barrow*, ed. William Whewell. Cambridge: Cambridge University Press.

Bergson, Henri. 1944. *Creative Evolution*. (English translation of *L'Évolution créatrice*, 1907). New York: Henry Holt and Company (originally published 1911).

Bernal, J.D. 1949. *The Freedom of Necessity*. London: Routledge and Kegan Paul.

Bohm, David. 1957. *Causality and Chance in Modern Physics*. Philadelphia: University of Pennsylvania Press.

Boltzmann, Ludwig. 1964. *Lectures on Gas Theory*. Trans. Stephen G. Brush. Berkeley/LA: University of California Press.

Bondi, Hermann. 1952. Relativity and Indeterminacy. *Nature* 169: 660.

Born, Max. 1949. *Natural Philosophy of Cause and Chance*. Oxford: Clarendon Press.

Bunge, Mario A. 1979. *Causality and Modern Science*. (3rd revised edition; originally published 1959, Harvard University Press). New York: Dover.

Burbury, S.H. 1894. *Nature*. 1308, 51 (November), 78.

Butterfield, Jeremy, and John Earman (eds.). 2007. *Handbook of the Philosophy of Science: Philosophy of physics*. Amsterdam: North-Holland.

Čapek, Milič. 1971. *Bergson and Modern Physics*. Dordrecht: D. Reidel.

Čapek, Milič, ed. 1976. *Concepts of Space and Time*. Boston: D. Reidel.

Clausius, Rudolf. 1867. *Abhandlungen über die mechanische Wärmetheorie*, vol. 2. Braunschweig: Vieweg.

Davies, Paul C.W. 1995. *About Time*. New York: Touchstone.

Dempsey, Liam. 2006. Written in the Flesh: Isaac Newton on the Mind–body Relation. *Studies in History and Philosophy of Science Part A* 37 (3): 420–441.

Dorato, Mauro. 2006. Absolute Becoming, Relational Becoming and the Arrow of Time: Some Non-conventional Remarks on the Relationship Between Physics and Metaphysics. *Studies in History and Philosophy of Modern Physics* 37: 559–576.

Earman, John. 1974. An Attempt to Add a Little Direction to 'The Problem of the Direction of Time'. *Philosophy of Science* 41 (1): 15–47.

Earman, John. 1986. *A Primer on Determinism*. Dordrecht: D. Reidel.

Earman, John. 1989. *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time*. Cambridge, MA: MIT Press.

Earman, John. 2006. The 'Past Hypothesis': Not even false. *Studies in History and Philosophy of Modern Physics* 37: 399–430.

Eddington, Arthur S. 1929. *The Nature of the Physical World*. New York: Macmillan.

Falk, Dan. 2016. A Debate Over the Physics of Time. *Quanta Magazine Newsletter*, July 19, 2016.

Feingold, Mordechai. 1990. Isaac Barrow: Divine, Scholar, Mathematician. In *Before Newton: The Life and Times of Isaac Barrow*, ed. Mordechai Feingold, 1–104. Cambridge: Cambridge University Press.

Feingold, Mordechai. 1993. Newton, Leibniz and Barrow Too: An Attempt at a Reinterpretation. *Isis* 84: 310–338.

Futch, Michael. 2008. *Leibniz's Metaphysics of Time and Space*. Dordrecht: Springer.

Gassendi, Pierre. 1649. *Animadversiones in Decimum Librum Diogenis Laërtii* (*Critical Remarks on the Tenth Book of Diogenes Laërtius*). Lyon: Guilliaume Barbier.

Gold, Thomas. 1962. The Arrow of Time. *American Journal of Physics* 30: 403–410.

Gold, Thomas. 1966. Cosmic Processes and the Nature of Time. In *Mind and Cosmos*, ed. R. Colodny, 311–329. Pittsburgh: University of Pittsburgh Press.

Greene, Brian. 1999. *The Elegant Universe*. New York: Random House.

Greene, Brian. 2004. *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*. New York: Alfred A. Knopf.

Grünbaum, Adolf. 1971. The Meaning of Time. In *Basic Issues in the Philosophy of Time*, ed. Eugene Freeman and Wilfrid Sellars, 195–228. La Salle: Open Court.

Hawking, Stephen W. 1988. *A Brief History of Time*. London: Bantam Press.

Herivel, John. 1965. *The Background to Newton's Principia*. Oxford: Clarendon Press.

Horwich, Paul. 1987. *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge, MA: MIT Press.

Huygens, Christiaan. 1986. *The Pendulum Clock, or Geometrical Demonstrations Concerning the Motion of Pendula as Applied to Clocks*. Trans. of *Horologium Oscillatorium* (1673) with notes by Richard J. Blackwell. Ames: Iowa State University Press.

Jones, Alexander. 2017. *A Portable Cosmos: Revealing the Antikythera Mechanism, Scientific Wonder of the Ancient World*. Oxford: Oxford University Press.

Kiefer, Claus. 2011. Time in Quantum Gravity. In *The Oxford Handbook of Philosophy of Time*, ed. Craig Callender, 491–502. Oxford: Oxford University Press.

Landels, John G. 1979. Water-Clocks and Time Measurement in Classical Antiquity. *Endeavour* 3 (1): 32–37.

Laplace, Pierre Simon de. 1812. *Théorie analytique des probabilités*. Paris: Ve. Courcier.

Leibniz, Gottfried Wilhelm. 1951. *Theodicy*. Trans. E.M. Huggard, London: Routledge and Kegan Paul; reprint edition, La Salle: Open Court, 1985.

Leibniz, Gottfried Wilhelm. 1969. *Philosophical Papers and Letters*. Ed. and trans. Leroy Loemker, 2nd ed, 1976. Dordrecht: D. Reidel.

Leibniz, Gottfried Wilhelm. 1981. *New Essays on Human Understanding*. Trans. Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press.

Leibniz, Gottfried Wilhelm. 1998. *Philosophical Texts*. Trans. and Ed. R.S. Woolhouse and Richard Francks. Oxford: Oxford University Press.

Leibniz, Gottfried Wilhelm. 2001. *The Labyrinth of the Continuum: Writings on the Continuum Problem, 1672–1686*. Ed., sel. and Trans. with an introductory essay, R.T.W. Arthur. New Haven: Yale University Press.

Loschmidt, J.Josef. 1876. Über den Zustand des Wärmegleichgewichtes eines Systemes von Körpern mit Rücksicht auf die Schwerkraft. *Sitzungsberichte der Akademie der Wissenschaften zu Wien, mathematisch-naturwissenschaftliche Klasse* 73: 128–142.

Mehlberg, Henryk. 1961. Physical Laws and Time's Arrow. In *Current Issues in Philosophy of Science*, ed. H. Feigl and G. Maxwell, 105–138. New York: Holt, Rinehart and Winston.

Morris, Richard. 1985. *Time's Arrows: Scientific Attitudes Toward Time*. New York: Simon and Schuster.

Musser, George. 2018. What is Spacetime? *Nature* 557, 10 May 2018.

Newton, Isaac. 1710. *Two Treatise of the Quadrature of Curves*. Trans. John Harris of *De quadratura curvarum*; reprinted in *The Mathematical Works of Isaac Newton*, vol. 1, ed. D.T. Whiteside. New York and London: Johnson Reprint, 1964.

Newton, Isaac. 1999. *The* Principia*: Mathematical Principles of Mathematical Philosophy*. Ed. and trans. I. Bernard Cohen and Anne Whitman. Berkeley: University of California Press.

Penrose, Roger. 2016. *Fashion, Faith and Fantasy in the New Physics of the Universe*. Princeton and Oxford: Princeton University Press.

Popper, Karl. 1956. The Arrow of Time. *Nature* 177, March 17.

Price, Huw. 1996. *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. New York/Oxford: Oxford University Press.

Price, Huw. 2011. The Flow of Time. In *The Oxford Handbook of Philosophy of Time*, 210–235, ed. Craig Callender. Oxford: Oxford University Press.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Sklar, Lawrence. 2015. Philosophy of Statistical Mechanics, *The Stanford Encyclopedia of Philosophy*, Fall 2015 ed, ed. Edward N. Zalta. https://plato.stanford.edu/archives/fall2015/entries/statphys-statmech/.

Smart, J.J.C. 1968. *Between Science and Philosophy*. New York: Random House.

Smith, George. 2008. "Newton's *Philosophiae Naturalis Principia Mathematica*", *The Stanford Encyclopedia of Philosophy*, Winter 2008 ed, ed. Edward N. Zalta. https://plato.stanford.edu/archives/win2008/entries/newton-principia/.

Smolin, Lee. 2013. *Time Reborn: From the Crisis in Physics to the Future of the Universe*. Toronto: Alfred A. Knopf.

Sobel, Dava. 1995. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York: Penguin.

Stein, Howard. 1968. On Einstein-Minkowski Space-Time. *Journal of Philosophy* 65 (1): 5–23.

Torretti, Roberto. 1999. On Relativity, Time-Reckoning, and the Topology of Time-Series. In *The Arguments of* Time, ed. J.N. Butterfield, 66–82. Oxford: Published for the British Academy by Oxford University Press.

Torretti, Roberto. 2007a. 'Rod contraction' and 'clock retardation': Two harmless misnomers? In *Filosofia, scienza e bioetica nel dibattito contemporaneo: Studi internazionali in onore di Evandro Agazzi*, ed. Fabio Minazzi, 1029–1039. Roma: Istituto Poligrafico e Zecca dello Stato, 2007.

Torretti, Roberto. 2007b. The Problem of Time's Arrow Historico-critically Reexamined. *Studies in History and Philosophy of Modern Physics* 38: 732–756.

Uffink, Jos. 2003. Irreversibility and the Second Law of Thermodynamics. In *Entropy*, ed. A. Greven, G. Keller, and G. Warnecke, 121–146. Princeton, NJ: Princeton University Press.

Uffink, Jos. 2007. Compendium of the Foundations of Statistical Physics. In ed. Butterfield and Earman, 924–1074.

Unger, Roberto Mangabeira, and Lee Smolin. 2015. *The Singular Universe and the Reality of Time*. Cambridge: Cambridge University Press.

Wheeler, John Archibald (with Kenneth Ford). 1998. *Geons, Black Holes and Quantum Foam*. New York/London: W. W. Norton.

Whitrow, G.J. 1961. *The Natural Philosophy of Time*. CA: Harper. London: Nelson.

Winnie, John. 1977. The Causal Theory of Space-time. In *Foundations of Space-Time Theories*, ed. John Earman, Clark Glymour, and John Stachel, 134–205. Wisconsin: Minnesota University Press.

Chapter 5
Special Relativity and the Lapse of Time

The objective status of becoming was strengthened rather than weakened by the special theory of relativity.
—Milič Čapek (1971, 233).

5.1 Introduction

If most people know one thing about Einstein, it is that he was the inventor of the theory of relativity. Many people have only a vague idea of what that theory is. In the popular press of his time it was presented as the theory that “everything is relative”. But those who understand a little more know that the main novel feature of Einstein’s theory of Special Relativity (hereafter SR) is the claim that *simultaneity is relative*. What is simultaneous with me-now when I am at rest will no longer be simultaneous with me-now if I am moving with great speed. In the vivid example of Roger Penrose, if two people are walking slowly past each other on Earth, the events each infers to be happening “now” on some planet in the Andromeda Galaxy (some 10^{19} km away) will be different by a few days from one another (Penrose 1989, 201).[1]

In short, there is, according to relativity theory, no unique set of events that are simultaneous with a given event, such as my looking at my watch at 3 p.m. That will depend on the state of motion of the reference frame from which we consider my looking at it. So there is no unique ‘now’, or world-at-an-instant. For many authors, that “seals the deal” for temporal becoming. Time cannot consist in the successive existence of objectively defined worlds-at-an-instant if there are no unique worlds-at-an-instant.

One of the first to make this argument explicit was Kurt Gödel in 1949, in a paper that is more famous for its argument for the possibility of time travel back into one’s own past. (We’ll come to that in due course, when we consider the implications of

---

[1] If each is walking at 4 km per hour in opposite directions, the time difference will be about 3 days. Similarly, if I get up from a sitting position and walk at about 3 km h^{-1} the difference will be about a day (Davies 1995, 70).

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_5

curved spacetime in Chap. 7.) But in the same paper Gödel argues that the relativity of simultaneity undermines the objectivity of the lapse of time, thus yielding "an unequivocal proof for the view of those idealistic philosophers who, like Parmenides, Kant and the modern idealists, deny the objectivity of change and consider change as an illusion or an appearance due to our special mode of perception" (Gödel 1949, 557).

Such a radical idealist view has not gained wide acceptance among modern philosophers of science. The dominant view is that although the amount of time that lapses between any two events will vary according to which reference frame is chosen, time lapse is nevertheless objective. What is refuted, according to these philosophers, is any notion of *objective becoming*. The events simply *are* at the locations in spacetime at which they occur. But there are certain pairs of events, the so-called spacelike separated events, that occur too far apart for light to be able to travel from one to the other; and such pairs of events can be made to appear in a different temporal order by varying the frame of reference from which they are viewed. That is, *a* can be before *b* in one inertial reference frame, *b* before *a* in another, and *a* and *b* simultaneous in a third. This fact, according to the dominant view, shows that no meaning can be ascribed to these events *becoming present* in any objective fashion. On this view, although the time lapse between events in spacelike separation is dependent on the inertial frame from which it is viewed, it is perfectly objective. Gödel demurred, holding that the very relativity of time lapse was proof of its lack of objectivity. But on neither view is there any objective becoming.

In this chapter I shall argue that these arguments mistakenly construe time lapse as what is measured by the time coordinate function associated with a given inertial reference frame. They fail to take into account that in relativity theory there is a bifurcation of the time concept into two distinct concepts: the *time-coordinate function* t, which tracks which events in a given frame are to be accorded as simultaneous with a given event; and the *proper time* τ, introduced by Hermann Minkowski in 1907, which tracks the rate of a process or the rate at which a thing ages as it proceeds along a particular path through spacetime. It is only the latter concept that is associated with the temporal becoming of events in succession, and therefore represents the time elapsed for such a process. The former, the t that is on a par with the three spatial coordinates x, y, and z, of an inertial reference frame, gives an objective representation of synchronicity, but one that is relative to frame. The coordinates assigned to a given event in one inertial frame are related to those in another by a set of formulas called the Lorentz transformations. But in the case of events that are further apart than can be connected by any process going at the speed of light—spacelike separated events—the differences in t-values in any such frame do not represent the time lapse between the events, since spacelike related events are not such as can occur in the same process. This bifurcation of time into two different time concepts t and τ is characteristic not only of SR, but of all relativistic physics, where every timelike curve represents the path of a possible process, whose rate of evolution is parametrized by proper time.

But in order to understand these concepts—inertial reference frame, the Lorentz transformation formulas, proper time, spacelike separation, timelike curves, and so forth—we need to take a step backwards, and consider the history of the idea of the relativity of motion, and the creation of the theory of Special Relativity by Albert Einstein and Hermann Minkowski.

5.2 The Origins of Relativity

The idea of the relative nature of motion was not Einstein's invention, nor did he claim it to be. It was advocated (with some qualifications) by René Descartes (1596–1650), and taken up in earnest by Christiaan Huygens (1629–1695), the doyen of mathematical physics in the seventeenth century prior to Newton and Leibniz. The following statement of Huygens contains a remarkably clear statement of the relativist position:

> It does not seem possible to understand what rest or motion there is in bodies except in respect of other bodies. For concerning motion we can imagine nothing but what changes the mutual distance and disposition of bodies among themselves. And so a body that moves is said to move in respect of other bodies with which it changes situation, and to rest in respect of those with which it conserves its situation. (*De motu corporum ex percussion. Appendice 1*; Huygens [1654] 1929, 111)

This is from a manuscript written by Huygens in 1654, "On the motion of bodies as a result of collision", although it remained unpublished till much later. But the same position informed Huygens' published work, and conditioned his response to Newton's discussion of absolute space and motion in the latter's masterwork, the *Principia* of 1687. For, Huygens said, if "motion is merely relative between two bodies", then the same relative motion can be produced by impressing motion on either one of them towards the other. "Indeed, absolutely the same effect results from either impression" (Stein 1977, 48). Such a relativity of motion, Huygens had come to believe, made Newton's assumption of an absolute space and time unnecessary. Huygens's championing of the mutuality and relativity of rectilinear motion made its mark on Newton's great rival, Gottfried Leibniz, who learned his physics under Huygens's tutelage when he was in Paris in the years 1672–1676.[2]

For his part, Newton vehemently rejected the idea that all motion could be relative. He believed there is a fact of the matter whether a given body has been set in motion with a given force, and that this would produce a change in the quantity of motion of that body as it moved in absolute space. If A is moved towards B—say, for example, if a ball is launched at some skittles at 10 km/h—then it is certainly true that, relatively to the ball, the skittles are moving towards it at the same speed. But no one has moved the skittles, whereas it took some force to launch the ball. The relative motion of the

---

[2] Here I am ignoring the changes in Huygens's unpublished views on relativity. He always held to the relativity of rectilinear motions, even in collisions. But by the time of Leibniz's sojourn in Paris (1672–6), he had come to believe that the centrifugal effect gave a criterion for motion in absolute space, as Newton would later argue. Huygens's subsequent re-espousal of the relativity of all motion was not the same as Leibniz's commitment to the Equivalence of Hypotheses concerning appearances, nor did he accept Leibniz's mature views that true motions could be identified by appeal to causes. See Mormino (1993) and (2011) for full discussion.

skittles to the ball, Newton believed, is the change of an “extrinsic denomination” in them, representing an accidental rather than an intrinsic change in the skittles themselves, and is not a true motion. By taking the motion as relative to this body or that one, he says, “every relative motion can be changed while the true motion is preserved, and can be preserved while the true one is changed, and thus true motion certainly does not consist in relations of this sort” (Newton 1999, 412).

Nevertheless, Newton acknowledges, “it is certainly very difficult to discover the true motions of individual bodies and actually to differentiate them from apparent ones, because the parts of that immovable space in which the bodies truly move make no impression on the senses” (1999, 414). He thinks the case is “not altogether hopeless”, as the apparent motions will be differences of true motions, and in some cases one can identify the causes producing the true motions. But, he acknowledges, if a whole system of bodies is moving inertially, that is, in a straight line with the same velocity, then the relative motions of these bodies will be the same as if the whole system is at rest. With customary precision he states this as follows:

> When bodies are enclosed in a given space, their motions in relation to one another are the same whether the space is at rest or whether it is moving uniformly straight forward without circular motion. (Corollary 5; Newton 1999, 423)

Newton proposed this as a corollary to his Laws of Motion—laws which he regarded as “accepted by mathematicians and confirmed by experiments of many kinds” (1999, 424). For by his second law, “a change in motion is proportional to the motive force impressed”, so that “from the sums and differences of motions tending in the same direction … there arise the collisions and impetuses with which the bodies strike one another” (1999, 423). The change in a body’s motion, $m\Delta v = m(v_1 - v_2)$, will be unaffected by setting the space in which they are enclosed moving with a velocity v_3, since adding v_3 to both v_1 and v_2 will leave $m\Delta v$ unaltered. For good measure, Newton adds that this is “proven by experience”: “on a ship, all the motions are the same with respect to one another whether the ship is at rest or moving uniformly straight forward” (423). This is a nod to Galileo’s lyrical thought experiment in his *Dialogue Concerning the Two Chief World Systems*:

> Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl with some fish in it … With the ship standing still, observe how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions … and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal … When you have observed all these things carefully, … have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. … The fish in the water will swim toward the front of their bowl with no more effort than toward the back … Finally the butterflies and flies will continue their flights indifferently towards every side … (Second Day; Galilei 1967, 186–87).[3]

---

[3] It should be noted that Galileo assumes all this to be taking place on the spherical surface of the Earth; for him, what requires no cause is the circular motion of a heavy body along the surface of the spherical Earth (assumed frictionless). It was in fact Descartes who proposed that the motion

In accordance with this, Newton acknowledged that it will not be possible to determine empirically whether a system of bodies in relative motion is at rest or moving inertially in a straight line. But he held that the very distinction between being at rest or moving inertially presupposes an absolute space with respect to which these motions are gauged.

The triumph of Newtonian physics brought with it an acceptance of Newton's philosophy of space and time, despite the difficulty of determining motion in absolute space. For example, one of those who did most to popularize Newton's physics in Europe was Leonhard Euler (1707–83). In 1748 he proposed that a body moving inertially will cover equal distances in equal times, so that Newton's First Law, the Law of Inertia, implicitly defines an ideal clock beating out absolute time (as we discussed above in Chap. 3). Also, by Corollary 5 a system of bodies moving inertially will preserve the same direction of motion in space and the same relative motions and positions as time passes. For Euler this was enough to support Newton's contention of the reality of space and time, notwithstanding the fact that the positions and absolute velocities of the bodies in space cannot be determined.[4]

But in the late nineteenth century the intellectual climate was very different, and dominated by an empiricist and positivist philosophy. Its most forceful advocate was the physiologist and philosopher Ernst Mach, who sought to cleanse scientific thinking by purging it of intruding metaphysical concepts. In his influential *Science of Mechanics* (1883), Mach analyzed how the concepts of mechanics arose in experience, with a heavy emphasis on how each mechanical quantity was related to the measurements that could be used to determine it. From this standpoint, Newton's "immovable" absolute space appeared to Mach as a symptom of "the influence of medieval philosophy, as though he had grown unfaithful to his resolve to investigate only actual facts" (Mach 1919, 223). According to Mach, all cosmological motions implicitly made reference to the fixed stars as a set of reference bodies relative to which they were moving, and to the rotation of the Earth as a time scale. Newton's admission that the parts of absolute space "make no impression on the senses" was tantamount to admitting that it was a dispensable construct. So were the ideas of a motion uniform in itself, and a time independent of change. Mach wrote:

> A motion may, with respect to another motion, be uniform. But the question whether a motion is *in itself* uniform, is senseless. With just as little justice, also, may we speak of an "absolute time"—*of a time independent of change*. This absolute time can be measured by comparison with no motion; it has therefore neither a practical nor a scientific value; and no one is justified in saying that he knows aught about it. It is an idle metaphysical conception. (Mach 1919, 224)

A similar line on absolute motion was taken by scientists under the influence of Kant, who were no more inclined than Mach to admit space and time as things existing independently of human experience. For them space and time were rather "forms of

---

that requires no cause is motion at a constant speed in a straight line. Newton's animus towards Descartes by the time he wrote the *Principia* was such that he sought to expunge all record of Descartes's (significant) influence on his own work.

[4] Euler (1748). See DiSalle (2009).

intuition” that we necessarily use in constructing our experiences. So by the time Einstein penned a popular essay on relativity for the London *Times* in 1919, he could say: “It has, of course, been known since the days of the ancient Greeks that in order to describe the movement of a body, a second body is needed to which the movement of the first is referred” (Einstein 1954, 229). Although his subsequent statement that the motion of a planet is described “by relation to the set of fixed stars” shows clear influence of Mach, the basic position is essentially a reprise of the position of Huygens quoted above, that all motion is relative to a body or bodies.

A major change in how the relativity of motion was conceived, however, is signalled by what he says next: “In physics the body to which events are spatially referred is called the coordinate system” (Einstein 1954, 229). Einstein explains this idea as follows:

> A coordinate system which is admitted in mechanics is called an “inertial system”. The state of motion of an inertial system is, according to mechanics, not determined uniquely by nature. On the contrary, the following definition holds good: a coordinate system that is moved uniformly and in a straight line relative to an inertial system is likewise an inertial system. By the “special principle of relativity” is meant the generalization of this definition to include any natural event whatever: thus, every universal law of nature which is valid in relation to a coordinate system C, must also be valid, as it stands, in relation to a coordinate system C', which is in uniform translatory motion relatively to C. (Einstein 1954, 229)

Here Einstein is collapsing together three concepts that we would now distinguish: (1) that of a coordinate system, which is a set of geometric axes such as the three orthogonal axes of space and one of time, $\{x, y, z, t\}$, together with an origin and scale; (2) a reference frame, which can be defined independently of the choice of a particular coordinate system; and (3) an inertial system, which is a system of bodies that are at rest relatively to one another, moving uniformly in a straight line with constant velocity. The idea is that if we have a reference frame defined by such an inertial system (an inertial reference frame), and a set of coordinates adapted to it, we have what Einstein is here calling a “coordinate system”.

But notice that there is a problem of circularity here: with respect to what are these bodies moving uniformly in a straight line? Newton’s answer, as we have seen, was that it is *with respect to absolute space* that the relative space defined by the bodies’ mutual situations is moving inertially (i.e. uniformly in a straight line, without circular motion). Mach had proposed that (since absolute space is unobservable) we implicitly take the system of fixed stars as our reference bodies. But that does not guarantee that the fixed stars themselves are moving inertially. What would show that they were not moving inertially, however, would be some unbalanced leftover accelerations in our descriptions of the motions of the planets, after we had supposed the centre of gravity of the whole solar system to be moving inertially with respect to the fixed stars.[5] This would show itself in relative displacements of bodies in a system we had supposed to be inertial.[6]

On this basis we can use the laws of mechanics to successfully identify an inertial system. This was recognized by certain theorists in the late nineteenth century. Building on an earlier proposal by Carl Neumann, the Leipzig physicist Ludwig Lange proposed the following construction. Suppose three free particles are projected from a common source in different directions along straight lines that are not all in the same plane, in such a way that their distances from one another remain mutually proportional. The three particles will then constitute an inertial system, and the content of the law of inertia is that any other free particle moving in a straight line relative to this system will also be moving inertially (in what Newton called a relative space). The fact that these particles are "free", of course, is guaranteed by appeal to the laws of motion, which guarantee that there are no unexplained accelerations—as would happen if the whole system were in fact rotating. It follows that there is therefore an equivalence class of relative spaces implicit in the laws of mechanics, in each of which bodies moving uniformly in straight lines without circular motion cover equal spaces in equal times; these are the inertial frames. A coordinate system adapted to such a frame is an inertial coordinate system.

Now, the laws of mechanics are invariant under a change of inertial coordinate system: if using these laws a given natural process is described relative to an inertial reference frame with coordinate system C, there will be an equivalent description in the coordinate system C' of an inertial frame that is in uniform translation relative to C. This is the effective content of Einstein's statement of the principle of relativity above. As we will see in Chap. 6, it can be expressed as the requirement that the laws of mechanics are invariant under a group of transformations, namely, boosts in uniform velocity v, without requiring reference to material rods and clocks that are supposed to be material embodiments of a coordinate system.[7] The effect of a boost

---

[5] As Robert DiSalle observes, "By the later 19th century, observations became sufficiently precise to reveal that there is in fact a leftover acceleration, namely the famous extra precession of Mercury. But that could not affect Newton's analysis in 1687". (DiSalle 2009, §2.7)—we will come to the precession of Mercury in Chap. 7 below.

[6] Intriguingly, this analysis is completely in accord with Leibniz's account, save for one important detail. Leibniz had proposed that we adopt as a fiction the hypothesis that a system of bodies (such as the fixed stars) maintain the same mutual situations over time. We then relate positions and changes of position to these bodies using the accepted laws of mechanics, and designate that hypothesis as true which gives the most intelligible description of the phenomena. Adventitious phenomena whose causes remained unexplained on the Ptolemaic hypothesis but explained on the Copernican, for instance, such as the retrograde motions of the planets, the motions of tides and apparent changes of position of distant stars, would serve to confirm that the Copernican hypothesis is the true hypothesis. But Leibniz thought, wrongly, that rotational motion was in itself (i.e. without reference to such phenomena) relative. See Chap. 7 and also Arthur (2013) for further discussion.

[7] Einstein himself states the content of the Special Theory of Relativity in the way I am suggesting in his later essay "Physics and Reality" (Einstein 1954, 308): "it is necessary to postulate invariance of all systems of physical equations which express general laws with respect to Lorentz transformations. The elaboration of this requirement forms the content of the special theory of relativity". In Appendix A I sketch a group theoretic derivation of special relativity using such a Langean conception of reference frames in preference to the rods and clocks in terms of which Einstein

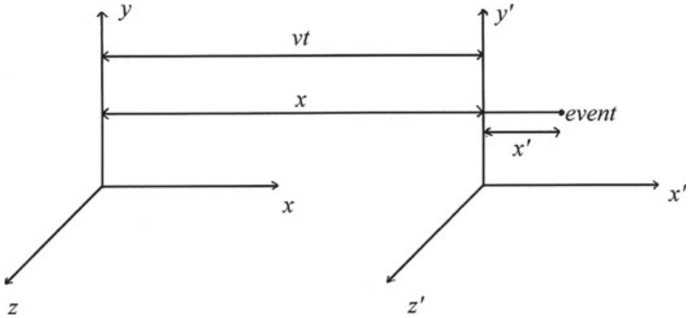


Fig. 5.1 A boost along the x-direction. The diagram illustrates a Galilean transformation, where the same (arbitrary) event is represented in two different coordinates system S and S' with S' moving at a velocity v along the x-axis relative to S, and with the same spatial origin at $t' = t = 0$

in velocity along the x-direction is that after a time t an event that was at position x will appear further to the left in the boosted system by a distance vt. The y and z co-ordinates will be left unchanged, and the time is assumed to lapse at the same rate in all the reference frames. So the classical transformation formulas constituting the relativity principle are

$$x' = x - vt, \;\; y' = y, \;\; z' = z$$

$$t' = t$$

These are known as the Galilean transformations, in honour of Galileo Galilei and his discussion quoted above. Any system of laws whose form does not change under a Galilean transformation, such as those constituting classical physics, is said to be Galilean-invariant (Fig. 5.1).

So understood, the relativity principle may be seen as a constraint on the behaviour of bodies implicit in the laws of physics. But it is a constraint that is wholly expected if, like Huygens and Mach, you regard motion as by definition relative to which other body or bodies you are taking to be at rest. For if only relative motions are observable, then boosting all such motions at any instant by a constant velocity v in a straight line in some direction will leave all the relative velocities between these

---

conceived them. The significance of Einstein's mistaken conception of inertial frames will become clear in Chap. 7, when we discuss Einstein's path to his General Theory of Relativity.

bodies unchanged.[8] For if two velocities $\underline{v}_1$ and $\underline{v}_2$ are each increased by the velocity $\underline{w}$, so that $\underline{v'}_1 = \underline{v}_1 + \underline{w}$ and $\underline{v'}_2 = \underline{v}_2 + \underline{w}$, then $\underline{v'}_1 - \underline{v'}_2 = \underline{v}_1 - \underline{v}_2$.

As Einstein realized, however, one of the crowning achievements of classical physics, Maxwell's electromagnetic theory, seemed not to be consistent with this constraint. This takes us to a consideration of the origins of Einstein's Special Theory of Relativity, and its reformulation by Minkowski into a theory of spacetime.

5.3 Special Relativity

According to Maxwell's electromagnetic theory, light propagates as a wave, consisting in oscillating electric and magnetic field disturbances, mutually generating one another as they go. It was one of the triumphs of Maxwell's theory, and a fact that helped to confirm it, that the speed of propagation of such an electromagnetic wave is exactly that of light, c,[9] the finite value of which had already been empirically established long before.[10] Just as water waves are propagated through the medium of water, it was supposed, so these electric and magnetic field disturbances must be motions in some underlying medium, the supposed electromagnetic *aether*.[11]

But what if you were moving relative to the light beam in the aether? If you were moving directly towards it at velocity v the light beam should appear to have a velocity of $v + c$, whereas if you were moving in the same direction as it at a velocity v it should have a velocity of $v - c$. Yet no such effect had been observed. As far as could be determined experimentally, the velocity of light in a vacuum is c,

---

[8] Here we may note that there is an ambiguity in the notion of the relativity of motion. It can mean that the motion of a body is always relative to whichever body is taken to be at rest; or it can mean only that all velocities are relative velocities—i.e. that there are no absolute, instantaneous velocities. But the latter type of relativity does not entail that all motions, including accelerations, are relative to a given reference system. As we shall see in Chap. 7, Einstein regarded the principle that all velocities are relative as yielding only a "restricted" theory of relativity, and sought to provide a properly general theory, in which all motions are relative to reference bodies, however moved.

[9] According to Maxwell's theory, these electromagnetic waves would have a velocity of $1/\surd(\varepsilon_0\mu_0)$, where $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m and $\mu_0 = 4\pi \times 10^{-7}$ H/m are the electric and magnetic constants appearing in his equations, giving a velocity of 2.998×10^8 m/s, the known velocity of light.

[10] Ole Rømer (1644–1710), a Danish astronomer working at the Royal Observatory in Paris, determined the velocity of light by timing the eclipses of Io, one of Jupiter's moons. He announced to the Observatory in a paper of 22 August 1676 that it took "about ten to eleven minutes" for light to traverse "a distance equal to the half-diameter of the terrestrial orbit", yielding a velocity of about 2.2×10^8 m/s, or about ¾ of the now accepted velocity of light. (His result was accepted by Huygens and Newton, but not fully ratified by astronomers until some decades later.)

[11] As emphasized by Roberto Torretti, however, H. A. Lorentz did not follow the British in conceiving electromagnetic waves as involving interaction between ponderable matter and the aether: for him the aether was motionless, in the sense that "no part of it is displaced with respect to its other parts, and that all perceptible motions of the heavenly bodies are motions relative to the aether" (Torretti 1999, 180; quoting Lorentz 1895, from his Collected Papers, vol. 5, p. 4). N.B. I follow Faraday, Maxwell and Torretti in using the British spelling 'aether', in order to distinguish this hypothesized medium from the chemical 'ether', $C_2H_5OC_2H_5$, a distinction lost with the American spelling.

and this is independent of any supposed motion relative to the aether permeating the vacuum—a result which, Einstein noted, is in conflict with the well known rule of addition of velocities in mechanics (Pais 2005, 139).

This is connected with a thought experiment Einstein had devised when he was sixteen years old. Suppose you are able to catch up with such a light beam by travelling at that speed c. Then all you would observe would be a stationary electric field, which, being stationary, would not be able to generate any magnetic field. But this is paradoxical, since light is by its nature electro-*magnetic*, and here there would be no magnetic component to generate an accompanying electric field disturbance, as Maxwell's theory says there should be. Light is either an electromagnetic wave in the aether or it is not: this should not depend on one's motion relative to it.

Physicists at the turn of the century (19th–20th) had found ingenious ways to accommodate Maxwell's theory to the apparent independence of light's velocity from motions in the aether. These culminated in the work of Joseph Larmor, Hendrik Antoon Lorentz and Henri Poincaré. They calculated that appearances would be preserved if (as first suggested by Fitzgerald) the length of an object moving in the aether underwent a contraction by a factor of $1/\gamma$, where $\gamma = 1/\surd(1 - v^2/c^2)$, with v its velocity relative to the aether, and c the velocity of light. Thus if a body's length at rest in the aether were l_0, then when moving through the aether with velocity v it would have a length of $l_0\surd(1 - v^2/c^2)$. For example, for a body moving at 60% of the speed of light, so that $v/c = 0.6$, this factor will be $\surd(1 - 0.36) = \surd(0.64) = 0.8$, so it would be shrunk to 80% of its rest-length in the direction of its motion relative to the aether. In addition, if material masses were increased by a factor γ, and if all forces, including molecular ones, varied with the velocity in the same way as electromagnetic ones, then motion with respect to the aether would be completely unobservable. In this way, Maxwell's equations would hold only in the frame of reference in which the aether is at rest, but everything would conspire in such a way as to make that frame unobservable. In particular, light would always appear to be moving at the speed c, even though this is only its speed in the aether. In order for this to work, Lorentz also had to adopt auxiliary *local times* for frames in motion with respect to the aether, but did not assign them any physical significance.[12] Joseph Larmor had shown (at least in the case of orbiting electrons), that these times would also be expanded or dilated by the same factor γ.

Under these assumptions, the coordinates of a system viewed from a frame of reference in inertial motion in the aether in the x–direction with local time coordinate t' will be transformed as:

$$x' = \gamma(x - vt), \ \ y' = y, \ \ z' = z$$

$$t' = \gamma(t - xv/c^2)$$

These transformations were first derived in this form by Poincaré (in June 1905) on the basis of Lorentz's theory, and ever after they have been known as the *Lorentz trans-*

---

[12] Lorentz defined his local time variable t' by $t' = t - (\mathbf{v} \cdot \mathbf{r})/c^2$ (Lorentz [1895] 1923).

formations (even though equivalent transformations had been derived previously by Larmor in a less elegant form).[13] Poincaré showed that these transformations entail that a system moving with velocity v in the aether, when viewed from an inertial frame moving at velocity w through the aether directly towards it, would have an apparent velocity of

$$V = \frac{(v + w)}{1 + vw/c^2}$$

It follows that a light ray moving at speed c would appear to have a velocity $(c + w)/(1 + cw/c^2)$, which reduces to c. This explains why light always appears to have the same velocity in a vacuum, and thus also accounts for the null result of experiments intended to establish the speed of the Earth through the aether, such as those of Michelson and Morley.

This classical theory thus has two components. One is constituted by the mathematical transformations that encapsulate the fact that motion relative to the luminiferous aether can never be observed. The other is a dynamical hypothesis to account for this fact, namely that physical objects undergo a length contraction and increase in mass as they move relative to the aether. Although there was no known mechanism for length contraction and mass increase, it seemed plausible at that time to suppose that intermolecular forces themselves, and everything constituting matter, were electromagnetic in nature.

Einstein, however, had undertaken a line of research that led him in a different direction. By early 1905, he had already learned how to account for the photoelectric effect by assuming that light was emitted in discrete pulses or packets of energy, light *quanta*. Strictly, the explanation of that effect only required that light be absorbed in such quanta, but Einstein boldly extrapolated it into a theory of emission too. This would, of course, have required a complete reworking of Maxwell's wave theory, which was, as it stood, incompatible with a theory of light as a stream of particles. Einstein was exploring whether the macroscopic behaviour of light could be deduced as a statistical effect of the combination of discrete processes at the microlevel. For reasons too complex to go into here, his particular attempts to supplant Maxwell's theory did not succeed. But he remained convinced that light was fundamentally particulate, its wave-form being a statistical effect at the macrolevel.

An immediate consequence of this perspective, however, is that it made the dynamical explanation of length contraction seem unconvincing. If light is not fundamentally a wave or disturbance in the aether, then there is no longer any explanation for phenomena such as length contraction occurring as a physical result of motion relative to the aether. One is left with the bare fact that light seems to have the same velocity independently of the velocity of its source relative to the observer. So if one

---

[13] As Poincaré wrote in an extended version of his 1905 note that was published in his (1906), "The reason why we can, without modifying any apparent phenomenon, confer to the whole system a common translation, is that the equations of an electromagnetic medium are not changed under certain transformations which I shall call the Lorentz transformations; two systems, one at rest, the other in translation, thus become exact images of one another".

takes this fact as a given, together with the principle of relativity, then something else has to give. And it seemed clear to Einstein that the only possibility for both these things to hold is if there is something askew with our classical concepts of space, time and velocity. If the length of a measuring rod is contracted by its relative motion, and this is not the result of the forces binding it together being altered by their motion relative to the aether, then perhaps it is a purely kinematical effect produced by distortions of space and time.

It is similar with the case of a magnet and a conductor in relative motion, the example with which Einstein begins his famous Special Relativity paper of 1905. The theory of Lorentz gives two different explanations of the current produced in the conductor, depending on whether the conductor is taken to be at rest in the aether with the magnet moving towards it, or the magnet at rest and the conductor moving. But the observed phenomena are precisely the same: an electric current of the same intensity is produced in the conductor in each case. Again, this suggested that the phenomenon is simply a result of their relative motion, and has nothing to do with the supposed aether.

So Einstein set about seeing whether he could provide a consistent theory by taking as postulates (1) the principle of relativity, according to which "the same laws of electrodynamics and optics will be valid for all frames of references for which the equations of mechanics hold good"; and (2) the so-called light postulate, namely "that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body" (Einstein 1905, 38). Einstein was further emboldened to undertake such a revision of the kinematics of motion by the influence of Mach's idea that the meaning of a physical concept is to be found in how one goes about measuring it. This he put into effect by having inertial frames (his "coordinate systems") constituted by measuring rods and material clocks:

> The theory to be developed is based—like all electrodynamics—on the kinematics of the rigid body, since the assertions of any such theory have to do with the relationships between rigid bodies (systems of co-ordinates), clocks, and electromagnetic processes. Insufficient consideration of this circumstance lies at the root of the difficulties which the electrodynamics of moving bodies at present encounters. (Einstein 1905, 38)

Possibly taking his cue from a reading of Poincaré's remarks on how one estimates the simultaneity of distant events,[14] Einstein proceeded to analyze this in terms of a light signal being sent from a source to a receiver, and then reflected back to the source. Assuming light always travels with the speed c in fact (and does not just *appear* to do so), the event of its reaching the receiver will be simultaneous with an event half way between the source's sending and receiving the signal, provided the two are relatively at rest. He then proceeded to show which events would be simultaneous with which if the source and receiver were in relative motion with a velocity v, and what effects this would have on any measurements of lengths and times in one system from an inertial frame in relative motion to it. By this means he

---

[14] Poincaré discussed the optical synchronization of clocks at rest in his (1898), pp. 371–384, and again in his (1904).

was able to derive the Lorentz transformations in the precise form that Poincaré had given.

In Einstein's theory, however, these equations now receive a very different interpretation. For now the time coordinate, "local time", does not represent a mere aid to calculation, as it did for Lorentz in his (1904); nor does it represent, as it did for Poincaré, a representation of the rate at which processes occurring in the aether in absolute time would *appear* to change when viewed from an inertial frame moving with respect to it. Instead, Einstein's time coordinates are all on a par: each gives the time of all those events that are simultaneous with a given event O from the perspective of a particular inertial coordinate system. But which set of events is simultaneous with O will vary according to the differing velocities of inertial systems relative to O. It follows, as Einstein concludes, that "we cannot attach any *absolute* signification to the concept of simultaneity". Two events that are simultaneous according to one inertial coordinate system "can no longer be looked upon as simultaneous when envisaged from a system which is in motion relatively to that system" (Einstein 1905, 42–43).

As a result, the *time dilation* that had previously been noted by Larmor and was implicit in the Lorentz transformations now took on a much more profound significance. The formula $t' = \gamma(t - xv/c^2)$, with $\gamma = 1/\surd(1 - v^2/c^2)$ and $x = vt$, may be rewritten as

$$t' = \gamma t(1 - v^2/c^2) = t\surd(1 - v^2/c^2) = t - \{1 - \surd(1 - v^2/c^2)\}t$$

This signifies, Einstein writes, that the time of a clock in a system moving with respect to an inertial coordinate system "is slow by $\{1 - \surd(1 - v^2/c^2)\}$ seconds per second" (49). This is usually summed up in the slogan "moving clocks run slow". But we must remember that this motion is a relative one: a clock moving relative to me will run slow, but in its own rest frame (the one relative to which it is stationary) it will tell the time just the same as usual. By the same token, from the point of view of someone in the rest frame of that clock, my clock will be running slow. This may appear paradoxical: how can both clocks be running slow? Faced with this clash with our intuitions, many people beat a retreat and talk of the clocks only *appearing* to run slow from each other's perspectives. In doing so they have failed to appreciate the radical nature of Einstein's proposal: the clocks *actually* run slow relatively to one another; and this is only possible if spatial and temporal intervals are not absolute, but relative to an inertial frame. Since this is hard to envisage in the abstract, we will return to it in Chap. 6, where we will demonstrate its consistency by means of some concrete, visualizable examples and straightforward mathematics.

We can describe the contrast between Einstein's theory and Lorentz's as follows. Where for Lorentz the unobservability of the motion of a light source in the ether was the result of its being precisely masked by the distortions of lengths and local times caused by this motion relative to the ether, Einstein took this unobservability of the velocity of light relative to the ether to indicate its unobservability in principle. He then promoted the indiscernibility of such a relative velocity from zero into an identity, turning what was for Lorentz a contingent fact into a necessary effect of this identity. His derivation of Special Relativity can thus be seen as an application of Leibniz's Principle of the Identity of Indiscernibles (PII): those states of affairs that are in principle indiscernible from one another are in fact identical. (In Chap. 7 we will see Einstein make a similar implicit application of the PII in deriving the Equivalence Principle of General Relativity.)

Hermann Minkowski was one of the first to appreciate the full implications of Einstein's theory for the concepts of space and time. In 1907 and 1908 he gave a sublime reconstruction of SR by reformulating it as a theory of a four-dimensional "world". This was achieved by the use of 4-vectors (vectors in 4 dimensions). Here the table had already been set for him by the work of Poincaré. The idea is a generalization of vectors in ordinary three-dimensional space, or 3-space. A 3-vector may be thought of as a line from the origin to some point (x, y, z). It will have a length given by $\surd(x^2 + y^2 + z^2)$. Already in his (1906) Poincaré had noticed that it is a consequence of the invariance of the speed of light that for any set of coordinates x, y, z, t, the expression $x^2 + y^2 + z^2 - c^2t^2$ is constant. Now if we substitute l for ict, where i is $\surd(-1)$, we have $x^2 + y^2 + z^2 + l^2$, which is the square of a length in a four-dimensional space (the quadratic form Q_P). On this basis, Poincaré was able to derive the Lorentz transformations by considering them as rotations in this four-dimensional space. But he was still interpreting Lorentz's local times as apparent times, as how absolute time would appear from moving frames.

For Minkowski the lesson of Einstein's relativity theory is that all of these Lorentzian local times are mutually equivalent: each gives a temporal projection onto a four-dimensional spacetime, what we call *Minkowski spacetime* and he called "the absolute world".[15] On the analogy with 3-vectors, we can think of an event $x = (ct, x, y, z)$ (called by Minkowski a *world point*) as defining a vector in a four-dimensional space M, a 4-vector. But this vector is a straight line in *spacetime*, rather than in space, whose length is $\surd(c^2t^2 - x^2 - y^2 - z^2)/c$. (The quadratic form of M assumed by Minkowski is $Q_M = c^2t^2 - x^2 - y^2 - z^2$, the negative of Poincaré's Q_P, since for any vector joining two spacetime points that can be spanned by a physical process, it is Q_M that is ≥ 0.)[16] Now any two events x and y can be joined by such a vector, but there will be three different kinds of such vectors, depending on the relative situation of the two events in question. (From now on we will simplify by holding y and z constant, and considering v as the velocity in the x-direction.)

1. If the two events could be the beginning and end of a light ray (in a vacuum), then $x = ct$, and so $\surd(c^2t^2 - x^2) = 0$. A vector joining them is called a *null-vector.*
2. If the two events are the beginning and end of an inertial motion from 0 to x, with a speed that is less than that of light ($w < c$), we will have $x = wt$. Then $c^2t^2 - x^2$

---

[15] In what follows I shall follow the terminology introduced by Minkowski in his 1908, although he had already articulated the mathematics and physics of his spacetime view in 1907. See (Minkowski 2012) for an English translation of these papers.

[16] This is because, as Minkowski relates, "at every worldpoint the expression $c^2dt^2 - dx^2 - dy^2 - dz^2$ is always positive, which is equivalent to saying that any velocity v is always less than c" (Minkowski 2012, 115).

$= c^2t^2 - w^2t^2$, which is necessarily greater than 0 because $w < c$. The length of the vector joining them is then $\surd(c^2t^2 - w^2t^2)/c$, or $t\surd(1 - w^2/c^2)$, which will be always positive. Such a vector is a *timelike* vector.

3. If two events are separated by an interval greater than can be traversed by any physical process going from one to the other, the vector joining them is a *spacelike* vector. For such pairs of events the Minkowski spacetime interval is imaginary, since $c^2t^2 < x^2$, so that $\surd(c^2t^2 - x^2) = iR$, where $i = \surd(c^2t^2 - x^2)$ and R is real-valued. Such a vector straddles a spacelike interval—spacelike in the sense that such pairs of events would correspond in the classical case (the limiting case where c is infinite) to events in the same instantaneous space, where no physical process could go from one to another. Accordingly, the length of a spacelike vector may be defined as $\surd(x^2 + y^2 + z^2 - c^2t^2)$, since for spacelike intervals it is now Q_P that is always positive.

Given this, we can imagine a set of timelike vectors joining all the events that can be reached by processes emanating from a given event O as origin: these events will all be "necessarily after" O, and can be said to constitute its *absolute future*. Correspondingly, all those events on timelike vectors that terminate in O are "necessarily before" O, and can be said to constitute its *absolute past*. Depicting this using two dimensions of space and one of time, we get a so-called Minkowski diagram of the "cone structure" of spacetime (Fig. 5.2):

Some things to note about this diagram: (1) since we have dropped one dimension, what are pictured as cones are in reality spheres; (2) this cone structure does not depict the shape of the whole of spacetime, but the structure existing at each and every point; (3) the time axis going through O is not unique, and may go through it at any angle between the vertical and a vector on the surface of the cone (which would be the path of light in a vacuum, a null vector); (4) each such time axis is associated with an inertial frame, and yields a different time coordinate for the set of events that are

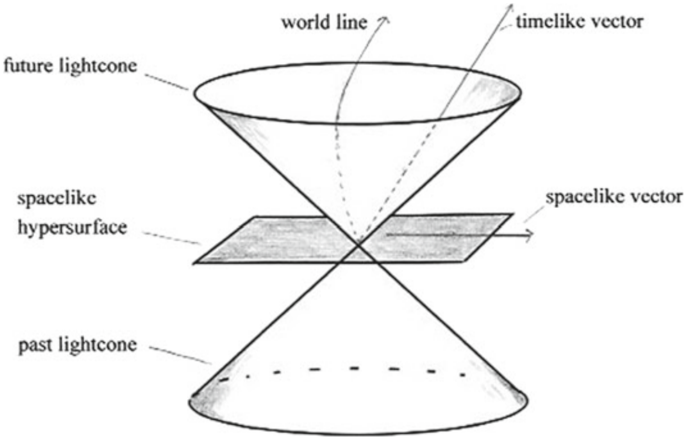


Fig. 5.2 The light cones of Minkowski spacetime

simultaneous with O in that inertial frame—in what Eddington picturesquely dubbed "the Elsewhere". So Einstein's relativity of simultaneity is depicted as the effect of taking different projections of time onto the whole of spacetime; (5) the diagram is conventionally drawn with the cones at 45° (i.e. π/4 radians). But if we took our units of scale as a metre and a second, we would have to flatten out the cones so that they made an angle of only 0.0000006° to the horizontal! In the limit where c goes to infinity, they would fuse into the horizontal hyperplane representing the unique present moment of absolute time.

So far this only treats inertial motions. What about non-inertial motions, accelerations and rotations? These will result in curved paths in spacetime. A curved path between two events that is everywhere timelike is called by Minkowski the *world-line* of a substantial point.[17] In order to treat such paths, Minkowski considered an infinitesimal neighbourhood of the point O, for which the timelike vector would be

$$d\tau = \frac{1}{c}\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}$$

By so doing he introduced (almost off-handedly) the extremely important concept of *proper time*, τ. To quote, "The integral $\tau = \int d\tau$ of this quantity, taken along the worldline from any fixed starting point P_0 to the variable endpoint P, we call the proper time of the substantial point at P".[18] This will give the time elapsed along the world-line from P_0 to P. This means that the proper time is specific to a particular path in spacetime along which it is integrated. In this respect it is completely different from coordinate time or the absolute time of classical physics. The beauty is, though, that given the Lorentz transformations, the integral of $d\tau$ along that curve will have the same value in an inertial system with coordinates (x, y, z, t) as in one with coordinates (x', y', z', t') related to (x, y, z, t) by those transformations. This means that proper time is an *invariant quantity*: it will have the same value as measured in any inertial frame. This fact will be of great significance for our discussion in this chapter.

A world line can be straight or crooked, but not so crooked that its velocity at any point could reach c (corresponding in our diagram to the tangent to the curve at that point being parallel to the surface of the cone). The world-line of a substantial point thus corresponds to the history of its path through spacetime. Still with one dimension suppressed, a three dimensional body is represented as a "world-tube".

Minkowski did not stop here, of course. He went on to show that the whole of mechanics and Maxwell's electromagnetic theory could be very elegantly recast using 4-vectors, thus bringing into sharp relief the full symmetry that Einstein had

---

[17] In his 1907 Minkowski called such a curve the "spacetime line"; he introduced the term "world-line" in his 1908. See Minkowski (2012, 98 and 112, resp.).

[18] (Minkowski 2012, 119) Actually, Minkowski had already introduced proper time in this way his lesser-known (1907). There he says the value of the integral of the element $d\tau$ "taken along the spacetime line from P_0 to a point P is called *proper time* (*Eigenzeit*), corresponding to the location of matter at the spacetime point P (This is a generalization of the concept of *local time* used by Lorentz (*Ortzeit*) in the case of uniform motion.)" (Minkowski 2012, 99).

perceived. From this new perspective, just as space and time are cast as projections of an invariant 4-dimensional spacetime, with each projection corresponding to the perspective of an inertial frame, so even mass and energy are projections of an invariant 4-vector "mass-energy", and electric and magnetic fields are projections in different inertial frames of 4-vectors representing electromagnetic fields. From this point of view, given all the misunderstandings of his theory as proving that "everything is relative", Einstein later regretted not having called his theory of relativity the theory of invariants (*Invariantstheorie*).[19]

5.4 Using Relativity Against Time Lapse

Now let us turn to the arguments that purport to show that the relativity of simultaneity in SR precludes the reality of becoming. We will begin with the argument of Kurt Gödel mentioned above. He argues that when two events *A* and *B* are separated by a spacelike interval the assertion that they are simultaneous "loses its objective meaning, insofar as another observer, with the same claim to correctness, can assert that *A* and *B* are not simultaneous (or that *B* happened before *A*)" (Gödel 1949, 557). This, he claims, allows one to prove that change is not objective, but "an illusion or an appearance due to our special mode of perception" (557). The proof he gives runs as follows:

> Change becomes possible only through the lapse of time. The existence of an objective lapse of time, however, means (or at least is equivalent to the fact) that reality consists in an infinity of layers of "now" which come into existence successively. But, if simultaneity is something relative in the sense just explained, reality cannot be split up into such layers in an objectively determined way. Each observer has his own set of "nows", and none of these various systems of layers can claim the prerogative of representing the objective lapse of time. (557–8)

We have seen that SR rules out the idea of a unique, absolute present: if the set of events that is simultaneous with a given event O depends upon the inertial reference frame chosen, and in fact is a completely different set of events (save for the given event O) for each choice of reference frame in inertial motion relative to the original, then there clearly is no such thing as *the* set of events happening at the same time as O. As Paul Davies writes (in a variant of the example given by Penrose above), if I stand up and walk across my room, the events happening "now" on some planet in the Andromeda Galaxy would differ by a whole day from those that would be happening "now" if I had stayed seated (Davies 1995, 70).

From these considerations Gödel concludes that time lapse loses all objective meaning. But from the same considerations Davies concludes, along with other modern philosophers of science, that it is not time lapse that should be abandoned, but the idea that events have to "become" in order to be real. "Unless you are a solipsist,"

---

[19] Isaacson notes Einstein's preference for the term "Theory of Invariants" (2007, 132), but without giving a reference.

he declares, “there is only one rational conclusion to draw from the relativity of simultaneity: events in the past and future have to be every bit as real as events in the present… To accommodate everybody’s nows, … events and moments have to exist all at once across a span of time” (Davies 1995, 71).

I contend that this is by no means a rational conclusion to draw. As I argued in Chap. 3 above, events “exist all at once” in a spacetime manifold only in the sense that we represent them all at once as belonging to the same manifold. But we represent them precisely as occurring at different times, or different spacetime locations, and if we did not, we would have denied temporal succession. The rational conclusion to draw, I submit, is that (according to Special Relativity) distant events that are simultaneous in some reference frame with a given event—for example, the event of my considering them—cannot be supposed to be ‘real’ or ‘existent’ for that event. They occur where and when they occur, not at the spacetime location from which I am considering them.

A similar scenario to that depicted by Penrose and Davies had already been described by Alfred North Whitehead in his *The Concept of Nature* (1920). He asks us to consider two perceivers, one on Earth and one (presumably in relative motion to him) on Mars:

> For the Earth-man there is one instantaneous space which is the instantaneous present, there are the past spaces and the future spaces. But the present space of the man on Mars cuts across the present space of the man on Earth. So that, of the event-particles which the Earth-man thinks of as happening now in the present, the man on Mars thinks that some are already past and are ancient history, that others are in the future, and others are in the immediate present. (Whitehead 1920, 177)

More elaborate arguments along the same lines have also been given independently by Hilary Putnam (1967) and by Wim Rietdijk (1966). Although the details of their arguments differ, both depend on the same scenario. We are asked to imagine two spatially distant inertial observers—let’s call them Alice and Bob—moving with respect to one another at an appreciable fraction of the speed of light. At a certain time according to Alice’s own inertial system, an event b that is happening to Bob is “present” or “now” for Alice, and we may imagine Alice’s being aware of this as the event a; but to Bob, the event happening to Alice that is simultaneous with b in his inertial system is not the event a, but another event p. Yet it is easy to set the relative velocity in such a way that p is in the future for Alice at the time that she is experiencing a. It follows that, if all those events are real which are present for a given observer in that observer’s inertial system, then b is real for Alice when she is experiencing a, and p is real for Bob when he is experiencing b. Thus if xRy denotes “x is real for y” we have bRa and pRb, so that, if R is transitive, then pRa (“p is real for a”) even though p is in the future for Alice when she is experiencing a. We are forced to conclude, reasons Putnam, “that future things (events) are already real” (Putnam 1967, 242), or as Rietdijk puts it, “that, being ‘past’ or ‘present’ for only one inertial system, an event can be shown to be determined in all other systems” (1966, 342), so that “there is determinism” and “there is no free will” (343).

Putnam, it should be said, acknowledges that simultaneity, although transitive within any given frame of reference, is not transitive between frames: “the relation ‘x

is simultaneous with *y* in the co-ordinate system of *x*' … is not transitive" (242–43). So he does not claim that all events exist "at once" in the sense of being mutually simultaneous. Nevertheless, he argues, the assumption that "all things that exist now according to my co-ordinate system are real", in combination with the principle that "there are no privileged observers", requires the relation *R* to be transitive (243). But if *R* is to be interpreted to mean that future events "already exist", as Putnam asserts, then this is to imply that they have, as of the earlier time, already occurred. A similar criticism applies to Rietdijk's conclusion: an event *p* can only be said to be "already 'past' for someone in our 'now'" (341) at location *a* in the sense that it has already occurred at *a*. But such a claim amounts to a denial of temporal succession.

In each case we are presented with an argument that begins with a premise that all events existing simultaneously with a given event exist (are real or are determined), and concludes that consequently all events in the manifold exist (are real or determined). But the conclusion only has the appearance of sustainability because of the equivocation analysed above in Chap. 3. If a point-event exists in the sense of occurring at the spacetime location at which it occurs, it cannot also have occurred earlier. But if the event only exists in the sense of existing in the manifold, then the conclusion that it *already* exists earlier—that such a future event is "every bit as real as events in the present" (Davies), or "already real" (Putnam)—cannot be sustained. Thus, far from undermining the notion of becoming, their argument should be taken rather to undermine their starting premise, that events simultaneous with another event are already real or already exist for it in a temporal sense. For to suppose that this is so, on the above analysis of their argument, inexorably leads to a conclusion that denies temporal succession.

This, in fact, was Gödel's point. As mentioned in the introduction to this chapter, he had already anticipated the objection that the relativity of time lapse "does not exclude that it is something objective". To this he countered that the lapse of time connotes "a change in the existing", and "the concept of existence cannot be relativized without destroying its meaning completely" (Gödel 1949, 558, n. 5). The dominant view, by contrast, would urge that the relativity of existence is avoided precisely by denying that time lapse constitutes a "change in the existing": the existence of events is their existence in a four-dimensional spacetime, and this does not change. As we saw in Chap. 3, however, the sense in which events and temporal relations "exist" in spacetime is not a temporal sense. So while it is true that there is no "change in the existing" of temporal relations among events, this does not mean that the events and relations *remain the same in time*, or that they are "already real" for a given event in their past. This would amount to a denial of the reality of temporal succession.[20]

What I wish to draw attention to here, however, is a premise that the dominant view shares with Gödel's: both assume that events are real or determined when they are present to an observer, with presentness construed in terms of simultaneity in the observer's frame of reference; i.e. they construe the reality of an event *in terms of the*

---

[20] See (Arthur 2006, 131–136). For similar analyses of the problematic notion of existence in a temporal context see the articles by Steve Savitt (2006) and Mauro Dorato (2006) in the same volume.

time co-ordinate function. Thus Putnam (1967) and Rietdijk (1966) each assume that becoming real or determined must occur relative to "an observer's inertial system", with time-lapse measured by the time co-ordinate function, as a premise in their *reductio* arguments against the reality of becoming real or determinate. The crucial premise here is the Gödelian one that for each individual observer, "the existence of an objective lapse of time … is equivalent to the fact that reality consists in an infinity of layers of 'now' which come into existence successively". That is, the time lapse between, for instance, two events in anyone's life history is given by the difference in the values of the time co-ordinate function in some particular inertial reference frame.

But this construal of time lapse in SR is false, as can be shown by an analysis of the much discussed Twin Paradox, devised by Paul Langevin in 1911. He asks us to imagine one twin staying at home while the other speeds off at a relative velocity which is an appreciable fraction of c, the speed of light, turns round, and returns at a similar velocity. When they are reunited, less time has elapsed for the travelling twin, who is consequently found to have aged less. But the discrepancy between the times elapsed for the two twins cannot be a discrepancy between times as measured by *co-ordinate time*—the time or "layer of 'now'" associated with some given inertial system—since in any one given inertial frame of reference the twins are apart for exactly the same time, as measured by the time co-ordinate of that frame. In any one such inertial frame, there is only one difference between the co-ordinates of these two points, and not one for each twin. In fact, however, the time taken for the twins to make each of their trips through spacetime from the point at which the travelling twin departed to the later point of their reunion must instead be determined by integrating the *proper time* along each twin's particular world line. When this is calculated for each of the two twins, it follows, as Langevin observes, that the twin who has aged less will be the one for whom "the time that has elapsed between the beginning and end of the path, the proper time" will be shorter, and this will be the twin whose motion is "furthest removed from uniform motion, the one most strongly accelerated" (Langevin 1911, 49; 1973, 296).[21] Thus the proper time for each twin is dependent on his or her path through spacetime; it is different because they take different paths through spacetime with different lengths. And the length of each path does not depend on the inertial coordinate system adopted.

So the root of the trouble with the "layer of now'" conception of time lapse is a failure to take into account the bifurcation of the classical time concept into two distinct time concepts in relativity theory. The time elapsed for each twin—the time during which they will have aged differently—is measured by the proper time along each path. The difference in the proper times for their journeys is not the same as the difference in the time co-ordinates of the two points in some inertial reference frame, since they each set off at some time t_1 and meet up at a time t_2 in any one

---

[21] It should not be thought, however, that it is the acceleration itself that produces the resulting dilation, as opposed to the difference in the length of the spacetime paths resulting from the fact that one twin has taken a path that is at least at some point non-inertial. In fact, Einstein himself was moved (in his 1918) to defend the consistency of special relativistic time dilation by appeal to its consistency with general relativistic time dilation, to be discussed in Chap. 7.

such coordinate system. Thus if time lapse were measured by such a time co-ordinate function, then both twins would be the same age. They are not. *Ergo*, time lapse (in the sense of how long a given process takes, how quickly it becomes) is not measured by the time co-ordinate function. So Gödel's "unequivocal proof" of the ideality of time falls flat on its face.[22]

It is puzzling that this simple consideration is not widely recognized; this suggests that there are other assumptions at work that mask its application. I believe they have to do with a *misconception of proper time as the time co-ordinate of the observer's rest frame*, and related misconceptions about *an observer* "*inhabiting an inertial frame*" and "*experiencing*" the events which are simultaneous with his or her state of consciousness. Rietdijk, for example talks of two spatially separated observers "experiencing the same present … in virtually the same inertial system" (Rietdijk 1966, 342), Grünbaum writes that for an organism M to experience an event at a time t is for it to be "*conceptually aware* of experiencing at that time either the event itself or another event simultaneous with it in M's reference frame" (Grünbaum 1976, 479), Putnam of "everything that is simultaneous to you-now in *your* co-ordinate system" being real, and Clifton and Hogarth (1995, 379) of two observers' "inhabit[ing] the same inertial frame". Although misconceptions about proper time are seldom stated explicitly, they also appear to be quite prevalent. Indeed, they afflict the understanding of SR itself, as witnessed by some of the attempted resolutions of the Twin Paradox.

These considerations will motivate us to take another look at the Twin Paradox in Chap. 6, in order to get clear on what is in an observer's (visual) experience in a relativistic context, and what is inferred; and to see more clearly how the distinction between proper time and co-ordinate time cleanly resolves the paradox without reference to the events one "experiences" as present undergoing a dramatic change, or implying that the discrepancy in the twin's ages is a General Relativistic effect. But for now I want to concentrate on the concept of proper time, and some of the misconceptions still attendant on it.

5.5 Proper Time and Proper Length

As I have argued, there are in fact two different measures of time in relativity theory: they have different formal measures, and different ontological baggage. This parallels the case for mass. In each case, what in classical physics had been thought to be a

---

[22] An anonymous referee to my (2008) objected that Gödel's argument depends only on the lapses of time being different for any two arbitrary curves connecting two timelike related events, and that Gödel does not assume that time lapse is measured by a time-coordinate function. But Gödel explicitly construes time lapse in terms of co-ordinate time in his argument from Special Relativity, where his argument against the "relativization of existence" crucially depends on this. This is supported by the interpretation of Yourgrau (1991), who construes Gödel's argument as depending on a conception of time lapse as relative to reference frame.

univocal or absolute property of the system turned out to be degenerate.[23] For in the transition to Einstein-Minkowski physics mass bifurcates into the relativistically invariant *proper mass* m_0, and the *relative mass* μ, or mass in an inertial frame in which it is moving at velocity v, whose quantity is a factor γ times the proper mass, so that $\mu = \gamma m_0 = m_0/\surd(1 - v^2/c^2)$ increases without limit as the relative velocity approaches c.

I believe much of the confusion about relativity comes from interpreting proper time as if it is simply the relative times of observers in their own rest frames. This misinterpretation is encouraged by the analogy with mass, but even more so, I will now suggest, by reading the case of time or *duration* as an exact analogue of that of space or *length*.[24] For the bifurcation of time in relativity theory is paralleled by a similar bifurcation in the concept of length. A body that is moving at a speed v with respect to a given inertial reference frame will, as already discussed, undergo a length contraction in the direction of its motion, so that its length $L = L_0/\gamma$, where L_0 is its *proper length*. The latter is defined as its length in the rest frame: if $v = 0$, $L = L_0$. Analogously, it may be thought, any periodic processes associated with the body will suffer a time dilation, so that $t = \gamma t_0$, with the result that in the rest frame where $v = 0$, $t = t_0$. *Proper time*, then, it may be asserted, *is just* t_0, *the time coordinate as measured in the rest frame*.

But this it is not! As we saw above, proper time was introduced by Hermann Minkowski in his famous 1908 paper (Lorentz et al. 1923, 73–91) as the integral along a given path in spacetime (a world-line) from any fixed starting point P_0 to the variable endpoint P of the quantity

$$d\tau = \surd(c^2dt^2 - dx^2 - dy^2 - dz^2)/c$$

As he proceeded to explain, x, y, z and t—the components of the vector OP, where O is the origin—are considered as functions of the proper time τ, and the first derivative of the components of this vector with respect to the proper time, $dx/d\tau$, $dy/d\tau$, $dz/d\tau$ and $dt/d\tau$, are the components of the *velocity 4-vector* $\mathbf{u}$ at P.

It is a consequence of this definition that the element of proper time $d\tau$ is not a complete differential. Arnold Sommerfeld, in his notes appended to Minkowski's 1908 paper when it was reprinted in a book (Lorentz et al. 1923, 92–96), remarked that Minkowski had mentioned this to him. He comments:

> [T]he element of proper time $d\tau$ is not a complete differential. Thus if we connect two world-points O and P by two different world-lines 1 and 2, then

---

[23] I use the term 'degenerate' here by analogy with quantum theory, where two or more different states may have the same energy level, but may nonetheless be distinguished by the application of an electric or magnetic field.

[24] I am indebted to Storrs McCall (private communication) for suggesting to me the relevance here of the analogy with proper length. I am also indebted to Kent Peacock for helping me eradicate some infelicities in my discussion of this in an earlier draft.

> $$\int_1 d\tau \neq \int_2 d\tau$$
>
> If 1 runs parallel to the t–axis, so that the first transition in the chosen system of reference signifies rest, it is evident that
>
> $$\int_1 d\tau = t, \int_2 d\tau < t$$
>
> On this depends the retardation of the moving clock compared with the clock at rest (94).

Evidently, Sommerfeld had already implicitly resolved the twin paradox in 1923 in essentially the same terms as I have given above—as had Langevin, as already noted.

What is crucial to this resolution is that the proper time calculated along a given path in spacetime is an *invariant quantity*: it retains the same value under transformation of inertial frame. It is for this reason that it "can claim the prerogative of representing the objective lapse of time", to use Gödel's own words (558), thus undermining his argument from the relativity of simultaneity to the unreality of time. Of course, Gödel assumed that an objective lapse would have to consist in a *global* plane of becoming, and therefore could not be relative to spacetime path; but, according to the point of view I am advocating here, this assumption is unwarranted in relativistic physics, where becoming is *local*, and dynamical change is parametrized by proper time, not co-ordinate time. It remains the case, of course, that the proper time is a maximum in the rest frame of an inertially moving object, and that in this circumstance it is numerically equal to the co-ordinate time t_0. For when $v = 0$, $\gamma = (1 - v^2/c^2)^{-1/2} = 1$, and $\tau = t_0$. But this is only numerical equality, not identity. It corresponds to the fact noted above that the *longest* time interval between two spacetime points in timelike separation is given by the straight line in spacetime connecting them. All other paths, whether the two inertial paths of the original Twin Paradox thought experiment or any other curved paths (as I shall argue further in Chap. 6), are shorter. But by the same token, Special Relativity is perfectly able to account for these non-inertial paths, and for each of them the proper time would be calculated by an integration along the path in question, not by the difference in time co-ordinates in any inertial frame. If proper time were the time co-ordinate in an inertial frame at rest, t_0, it would not be applicable to such curved paths. In contrast, proper *length* is the length of an object—a metre stick, say—in a specific frame of reference, namely, the inertial frame in which it is at rest.

Still, it may be objected, proper length is nevertheless also an invariant quantity. Just as the length of a path joining two events in timelike separation is invariant under change of frame, so is the length of a curve joining two events in spacelike separation. Indeed, it is often argued that the analogy between it and proper time is perfect: "proper length is the invariant interval of a spacelike path whereas proper time is the invariant interval of a timelike path".[25] Thus, it is suggested, the definition of proper length should be generalized so that it is the exact analogue of proper time: a line integral along a curve joining two spacelike separated events. But a little further reflection shows that this cannot be right: an arbitrary curve joining two spacelike separated events is not generally a length of some object. It is only the length of an object if all the points on the curve are simultaneous in some given reference frame. And while the path integral along such a curve is indeed independent of the choice of reference frame, it does not, despite its name, represent a *path* in the usual sense of that term. In the usual sense, a path is a series of positions that can be successively traversed—as, for instance, by Harvey Brown's waywiser (Brown 2005, front cover, p. 8)—and such an interval is timelike. If "path" is taken to mean a possible physical trajectory, then there are no "spacelike paths" in Minkowski spacetime. Proper length is correctly defined as the path integral, not along *an arbitrary curve* joining the endpoints of the metre stick at the same time, but along *the shortest curve*, which is a straight line joining them in the frame at which the stick is at rest. If (elapsed) proper time were the strict analogue of this, it would be the longest time between two timelike separated events, which would be the time in a frame of reference at rest, i.e. the co-ordinate time in that frame. But it is not. The proper time is defined along any timelike curve joining them, and is independent of the inertial frame in which it is evaluated, whereas proper length, because it is the interval between two events *at the same co-ordinate time*, is specific to a particular reference frame.

Thus *proper time* has a fundamentally different character from proper length. Although both are invariant under change of frame, proper length is the length of an object in its own rest frame, whereas proper time is independent of frame. In this respect proper length is analogous to proper mass. (It differs from the latter, however, in that proper mass seems to be an essential characteristic of an elementary body (such as an electron), whereas proper length is a contingent one.) At any rate, there is a fundamental dissymmetry between duration and length in Special Relativity, somewhat obscured by talk of their embodiments in observers' clocks and rods. For whereas an observer's clock measures proper time elapsed along a path, a dynamical variable specifiable independently of reference frame, the proper length of the observer's measuring rod is specific to the inertial frame in which the observer is at rest. So proper time is not analogous to proper length, and as an invariant dynamical variable it assumes an importance in relativistic physics that the latter simply does not possess. Thus, ironically, there is a sense in which Minkowski's introduction of proper time undermines his famous pronouncement at the beginning of his paper about the demise of time: "Henceforth space by itself, and time by itself, are doomed

---

[25] Quoted from an article on proper length in Wikipedia (http://en.wikipedia.org/wiki/Proper_length: May 5, 2007). The author suggested a generalization of proper length so that it is given by the line integral $L = c \int_P \sqrt{[-g_{\mu\nu} dx^\mu dx^v]}$, where $g_{\mu\nu}$ is the metric tensor for the spacetime with + --- signature, normalized to return a time. Since I published this criticism in (Arthur 2008), this Wikipedia article has (as of March 15, 2016) been emended, so that the latter formula is now correctly stated to represent the invariant proper distance along a path, and is correctly distinguished from proper length, "the length of an object in the object's rest frame".

to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality" (Minkowski 1908, 75).[26]

I am by no means the first to point out the radical implications of relativity theory for our understanding of time. As Milič Čapek has stressed in several publications (1966, 1975, 1976), the invariance of Minkowski's relations of being in the absolute past or future of an event means that in relativity theory the role of time is strengthened and made more distinct than in classical physics. The distinction between proper time and coordinate time is stressed by Larry Sklar in his treatment of the twin paradox[27]; and Peacock (1992), has also discussed the paradox in terms of a comparison between the proper times of the twins while they are spatially distant. But perhaps the clearest explanation of the distinction between "time co-ordinate" and "proper time" and its significance was given by Howard Stein:

> Proper time is not a quantity attached to space-time points or to pairs of space-time points; it is in this respect a notion utterly different from the quantity "time" or "time interval" of pre-relativistic theory … The fundamental physical role of proper time comes from the principle (here stated roughly) that whenever a process takes place along a well-defined line of space-time ("world-line"), the time rates in the dynamical principles that govern that process are to be understood in terms of the proper time along that line (and *not* in terms of a "time coordinate"…)[28]

Yet it seems to me that the significance of this degeneracy of time brought out by relativity theory is still largely unrecognized. Philosophers and physicists continue to write as if it is the time co-ordinate function, or time in relation to an inertial observer, and not proper time, that measures the duration of processes in relativistic physics. This is implicit in all discussions that agree with Gödel in construing the objective lapse of time in terms of an infinity of layers of "now", with these planes of simultaneity picked out by the time co-ordinate function in an inertial reference frame, such as the arguments of Davies, Putnam and Rietdijk discussed above.[29]

---

[26] Minkowski's judgement is echoed by Einstein in his essay "The Problem of Space, Ether and the Field in Physics": "Hitherto it had been silently assumed that the four-dimensional continuum of events could be split up into time and space in an objective manner… With the discovery of the relativity of simultaneity, space and time were merged in a single continuum … " (1954, 281–82).

[27] Sklar (1974, 268) correctly points out that, whereas "'co-ordinate time' between two events is relative to a given inertial frame", "[p]roper time is defined only for events at timelike separations and only relative to a particular spacetime curve between the events. On the other hand it is an invariant notion".

[28] (Stein 1968, 11, fn. 6). This quotation from Stein was my starting point for the line of argument developed here. Cf. also p. 16: "… 'a time co-ordinate' is not 'time.' Neither *a* nor *b* is, in any physically significant sense, 'present' (or past) for any observer at *c*—regardless of his velocity—for neither has already become for *c* (nor has *c* for them); but *a* has already become for *b*, and can influence it" [Here *a* and *b* are connectible by a timelike vector *ab*, the other pairs by spacelike vectors *ac* and *bc*.].

[29] A particularly striking example is provided by Vesselin Petkov, who, despite clearly recognizing the distinction between Einstein's relative time and Minkowski's proper time, regards the view that time flows as "unscientific". His argument is that if time really flowed then one would have to be able to pick out a unique (global) present moment, "which is the central element of the concept of time flow" (Petkov 2012, 31). We agree that this decisively rules out the classical conception of

Let me close with a neat argument devised by Steven Savitt that links these considerations from relativity theory back to the stock objection (considered above in Chap. 4) that the idea of time flowing at a certain rate is nonsensical.[30]

Savitt notes that in the Minkowskian treatment of relativity in terms of 4-vectors there is a 4-vector **u** representing a generalized notion of velocity. This gives a four dimensional representation of how fast a particular segment of a world-line is being traversed. It is defined by

$$\mathbf{u} = (cdt/d\tau, dx/d\tau, dy/d\tau, dz/d\tau)$$

Its length is therefore given by

$$\surd(\mathbf{u}\cdot\mathbf{u}) = \surd\left\{(cdt/d\tau)^2 - (dx/d\tau)^2 - (dy/d\tau)^2 - (dz/d\tau)^2\right\}$$

But by the definition of $d\tau$ as $= \surd(c^2dt^2 - dx^2 - dy^2 - dz^2)/c$, this gives

$$|\mathbf{u}| = \surd(\mathbf{u}\cdot\mathbf{u}) = \surd\left\{(d\tau/d\tau)^2\right\}c = c$$

This tells us that the instantaneous rate of change of some process occurring along a world-line with respect to the proper time along the curve is invariant: it is always the same, no matter what inertial reference system is chosen. Relativity theorists customarily choose units of spacetime so that $c = 1$, in which case $|\mathbf{u}|$, the rate at which a four dimensional path is being traversed, is always 1. Any path that traverses some space (according to a given reference frame) will result in at least one of the three components $dx/d\tau$, $dy/d\tau$, $dz/d\tau$ being greater than zero, meaning that the time component $cdt/d\tau$, will be correspondingly reduced. The greater the spatial distance covered by a process from one event to another, the shorter the time of the process. As Brian Greene has expressed this, "clocks that move through space in different ways tick off time at different rates (because they divert different amounts of their motion through time into motion through space)" (Greene 2004, 234). If the path is that of a system (say, an observer, "Bob") remaining at rest in the inertial frame, however, the spatial component of the 4-velocity will be identically zero, yet the observer's path will still be traversed at the same rate, $|\mathbf{u}|$. But now, as Savitt notes, "there is only one dimension left in which he can have this speed, the first or temporal dimension".[31] Even an observer who is stationary in some frame of reference still traverses a world-line through spacetime at the same 4-velocity, $|\mathbf{u}|$, and this is now c times the rate of passage of time, $dt/d\tau$. For such an observer, time flows at the rate of one second of coordinate time per second of proper time.

---

becoming in terms of a world-wide now; but he fails to acknowledge that this is not injurious to local becoming, where time lapse is tracked by proper time.

[30] Savitt presents the following argument in an unpublished paper from 2004, for the use of which I thank him. He begins the argument in (Savitt 2009, 355–6) but then concludes instead (360–1) with an argument of David Mermin's (Mermin 2005, 86), deriving from (Greene 1999, 47–51).

[31] The quotation is from Savitt's unpublished paper of 2004 (see previous footnote).

Perhaps this substantiates the claim that, even without motion, “time lapses with an even tenor”, as Gassendi, Barrow and Newton had claimed. At any rate, it shows that we can give sense to the idea of time flowing at a certain rate (in the sense I am defending here, with the moving now rejected), since we can after all make sense of the idea of different rates of flow along different time-like paths through spacetime.

5.6 Summary

- In this chapter I presented a short history of the idea of the relativity of motion, showing how it developed out of the ideas of Galileo, Huygens, Newton and Mach. This was followed by an account of Einstein’s development of the special relativity theory through his critiques of the classical theories of electromagnetism of Lorentz and others, and the seminal contributions of Poincaré and especially Minkowski in presenting this in a four-dimensional version.
- I have argued that time is a degenerate concept, which bifurcates into two different concepts in relativity theory. *Co-ordinate time* is used to track the synchrony of distant events, but it no longer has the classical role of tracking a worldwide hyperplane of becoming, as it did in classical theory. Instead it is *proper time* that measures time elapsed, and thus gives the true measure of the duration and rate of processes in spacetime.
- This bifurcation of time’s roles is masked by a tendency to assimilate proper time to time in an observer’s rest frame, by analogy with proper length, a tendency which is encouraged by unwarranted talk of an observer “inhabiting an inertial frame” and “experiencing” the events which are simultaneous with his or her state of consciousness. I give critiques of various erroneous attempts to refute the possibility of time flow on this basis.
- But whereas proper length is specific to a rest frame, proper time is not; its intervals are path-dependent, frame-independent, and invariant under change of reference frame. It is this proper time that measures the time elapsed for travellers in spacetime, which consideration is sufficient to resolve the Twin Paradox.
- Finally I argue that the four-dimensional view pioneered by Minkowski licenses the conclusion that time passes even for observers at rest in some inertial frame.

References

Arthur, Richard T.W. 2006. Minkowski Spacetime and the Dimensions of the Present. In ed. Dieks, 129–155.

Arthur, Richard T.W. 2008. Time Lapse and the Degeneracy of Time: Gödel, Proper Time and Becoming in Relativity Theory. In *The Ontology of Spacetime II*, ed. Dieks, 207–228. Amsterdam: Elsevier.

Arthur, Richard T.W. 2013. Leibniz’s Mechanical Principles (c. 1676): Commentary and Translation. *Leibniz Review* 23: 101–116.

Brown, Harvey R. 2005. *Physical Relativity: Spacetime Structure from a Dynamical Perspective*. Oxford: Clarendon Press.

Čapek, Milič. 1966. Time in Relativity Theory: Arguments for a Philosophy of Becoming. In *Voices of Time*, ed. J.T. Fraser, 434–454. New York: Brazilier.

Čapek, Milič. 1971. *Bergson and Modern Physics*. Dordrecht: D. Reidel.

Čapek, Milič. 1975. Relativity and the Status of Becoming. *Foundations of Physics* 5 (4): 607–617.

Čapek, Milič (ed.). 1976. *Concepts of Space and Time*. Boston: D. Reidel.

Clifton, Rob, and Mark Hogarth. 1995. The Definability of Objective Becoming in Minkowski Spacetime. *Synthese* 103: 355–387.

Davies, Paul C.W. 1995. *About Time*. New York: Touchstone.

DiSalle, Robert. 2009. Space and Time: Inertial Frames. In *The Stanford Encyclopedia of Philosophy*, Winter 2009 ed, ed. Edward N. Zalta. http://plato.stanford.edu/archives/win2009/entries/spacetime-iframes/.

Dorato, Mauro. 2006. The Irrelevance of the Presentist/Eternalist Debate for the Ontology of Minkowski Spacetime. In ed. Dieks, 93–109.

Einstein, Albert. 1905. Zur Elektrodynamik bewegter Körper. *Annalen der Physik* 17 (1): 891–921. (Trans. Lorentz et al. 1923, 35–65).

Einstein, Albert. 1918. Dialog über Einwände gegen die Relativitätstheorie. *Die Naturwissen–schaften* 48, 697–702 (29 November; an English translation exists on the web at http://en.wikisource.org/wiki/Dialog_about_objections_against_the_theory_of_relativity).

Einstein, Albert. 1954. *Ideas and Opinions*. New York: Crown Publishers.

Euler, Leonhard. 1748. Réflexions sur l'espace et le temps. *Histoire de l'Academie Royale des sciences et belles lettres* 4: 324.

Galilei, Galileo. 1967. *Dialogue Concerning the Two Chief World Systems*. Trans. S. Drake. Berkeley: University of California Press.

Gödel, Kurt. 1949. A Remark about the Relationship between Relativity Theory and Idealistic Philosophy. In *Albert Einstein: Philosopher-Scientist*, ed. Paul Arthur Schilpp, 557–562. New York: Tudor.

Greene, Brian. 1999. *The Elegant Universe*. New York: Random House.

Greene, Brian. 2004. *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*. New York: Alfred A. Knopf.

Grünbaum, Adolf. 1976. The Exclusion of Becoming from the Physical World. In ed. Čapek, 471–499.

Huygens, Christiaan. [1654] 1929. *De motu corporum ex percussion. Appendice*; *Oeuvres Complètes de Christiaan Huygens, XVI.* The Hague: Martinus Nijhoff.

Isaacson, Walter. 2007. *Einstein: His Life and Universe*. New York: Simon and Schuster.

Langevin, Paul. 1911. L'evolution de l'espace et du temps. *Scientia* 10: 31–54.

Langevin, Paul. 1973. The Evolution of Space and Time. *AMS Historica* 108: 285–300. (Trans. J. B. Sykes and P. Langevin, 1911).

Lorentz, H.A. 1904. Electromagnetic phenomena in a system moving with any velocity smaller than light. In *Royal Academy of Amsterdam, Proceedings 6*, 809–831 (Reprinted without the last section in Lorentz et al. 1923, 9–34).

Lorentz, H.A., Einstein, A., Minkowski, H., and Weyl, H. 1923. *The Principle of Relativity*. Methuen (Reprinted, Dover 1952).

Mach, Ernst. 1883. *Die Mechanik in ihrer Entwickelung, historisch-kritisch dargestellt*, 2nd ed. Leipzig: Brockhaus.

Mach, Ernst. 1919. *The Science of Mechanics*. Trans. Thomas J. McCormack (Mach 1888), 4th ed. Chicago/London: Open Court.

Mermin, N.David. 2005. *It's About Time: Understanding Einstein's Relativity*. Princeton and Oxford: Princeton University Press.

Minkowski, Hermann. 1907. *The Fundamental Equations for Electromagnetic Processes in Moving Bodies*. Trans. Vesselin Petkov, 51–110 (Minkowski 2012).

Minkowski, Hermann. 1908. *Space and Time*. Trans. Vesselin Petkov, 111–125 (Minkowski 2012).

Minkowski, Hermann. 2012. *Space and Time: Minkowski's Papers on Relativity*. In ed. Vesselin Petkov. Montreal: Minkowski Institute Press.

Mormino, Gianfranco. 1993. *Penetralia motus. La fondazione relativistica della meccanica in Christiaan Huygens, con l'edizione del Codex Hugeniorum 7A*. Firenze: La Nuova Italia.

Mormino, Gianfranco. 2011. Leibniz entre Huygens et Newton: force centrifuge et relativité du mouvement dans les lettres de 1694. In *Natur und Subjekt. IX. Internationaler Leibniz- Kongress Vorträge*, vol. 2, ed. Herbert Breger, Jurgen Herbst, and Sven Erdner, 697–705. Hannover: Druckerei Hartmann.

Newton, Isaac. 1999. *The Principia: Mathematical Principles of Mathematical Philosophy*. Trans. and ed. I. Bernard Cohen and Anne Whitman. Berkeley: University of California Press.

Pais, Abraham. 2005. *Subtle is the Lord: The science and the life of Albert Einstein*. Oxford: Oxford University Press (Reissued with a new foreword by Sir Roger Penrose, 1982).

Peacock, Kent A. 1992. A New Look at Simultaneity. In *Philosophy of Science Association 1992*, vol. I, ed. D. Hull, M. Forbes, and K. Okruhlik, 542–552. East Lansing: Philosophy of Science Association.

Penrose, Roger. 1989. *The Emperor's New Mind*. Oxford: Oxford University Press.

Petkov, Vesselin. 2012. *Editor's Introduction to (Minkowski 2012)*, 1–37.

Poincaré, Henri. 1898. La mesure du temps. *Revue de métaphysique et de morale* 6: 371–384.

Poincaré, Henri. 1904. L'état actuel et l'avenir de la physique mathématique. *Bulletin des sciences mathématiques* 28: 302–324.

Poincaré, Henri. 1906. Sur la dynamique de l'electron. In *Rendiconti del Circolo matematico di Palermo*; also in Poincaré, Oeuvres (Paris, 1954), vol. 9, 494–550.

Putnam, Hilary. 1967. Time and Physical Geometry. *Journal of Philosophy* 64 (8): 240–247.

Rietdijk, C.W. 1966. A Rigorous Proof of Determinism Derived from the Special Theory of Relativity. *Philosophy of Science* xxxiii (4): 341–344.

Savitt, Steven F. 2006. Presentism and Eternalism in Perspective. In ed. Dieks, 111–127.

Savitt, Steven F. 2009. The Transient Nows. In *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony*, ed. W. Myrvold and J. Christian, 339–352. Berlin: Springer.

Sklar, Lawrence. 1974. *Space, Time, and Spacetime*. Berkeley: University of California Press.

Stein, Howard. 1968. "On Einstein-Minkowski Space-Time". *Journal of Philosophy* 65 (1): 5–23.

Stein, Howard. 1977. Some Philosophical Prehistory of General Relativity. In ed. Earman, Glymour and Stachel, 3–49.

Torretti, Roberto. 1999. *The Philosophy of Physics*. Cambridge: Cambridge University Press.

Whitehead, Alfred North. 1920. *The Concept of Nature*. Cambridge: Cambridge University Press (Reissued Mineola, NY: Dover Publications, Inc., 2004).

Yourgrau, Palle. 1991. *The Disappearance of Time: Kurt Gödel and the Idealistic Tradition in Philosophy*. Cambridge: Cambridge University Press.

Chapter 6
Relativity and the Present

> *The view that now prevails, since Einstein's theory has come to be regarded as an essential part of physics, is that time is an aspect of the universe which depends on the observer.*
>
> —Whitrow, *Time in History* (1988, 173).

6.1 Introduction

According to a post in the Associated Press in early 2017, "Scientists have detected a black hole that's taken a record-breaking decade to devour a star—and it's still chewing away. The food fest is happening in a small galaxy 1.8 billion light-years from Earth."[1]—except, of course, that this last sentence is self-contradictory. If the light that astronomers are seeing is from a galaxy 1.8 billion light-years from Earth, then the event described took place 1.8 billion years ago, so to say that the event "is happening" is a stretch, to say the least, as is the statement that it is "still" chewing away. In the same vein, someone wrote to an online site wondering how it could be possible that we can see the Crab Nebula some 6500 light years away, given that Chinese astronomers saw the supernova that gave birth to it less than 1000 years ago—"meaning that light from [that event] shouldn't have reached us yet for us to see it, right?".[2] No, wrong: the light from the supernova would have taken some 6500 years to reach the Chinese astronomers too, so that the supernova would have occurred some 7500 years ago.

What these examples show is just how difficult it is to rid ourselves of the idea of the world existing at an instant. We are aware of the things around us, including the planets and stars as we see them, and for us this constitutes "the universe". But even allowing for the fact that what we see at cosmic distances happened long ago, we still imagine that there is a fact of the matter about what is happening *now*. In classical physics it was assumed that the world-at-an-instant is a well-defined concept. Classi-

---

[1] "Black hole taking more than a decade to gobble up star", *The Associated Press*, posted Feb 06, 2017 4:22 PM ET.

[2] https://www.quora.com/, accessed August 29, 2017.

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_6

cal spacetime could be thought of as a succession of three-dimensional instantaneous spaces threaded together along the time dimension, constituting a four-dimensional structure, absolute spacetime[3]—at least it could in retrospect, after Minkowski had introduced the notion of spacetime in his reformulation of Einstein's relativity theory. But as we saw in the previous chapter, in the Minkowski spacetime of SR there is no unique world-at-an-instant, because simultaneity is relative to the inertial frame chosen. In the words of Carol Rovelli, "The idea that a well-defined *now* exists throughout the universe is an illusion, an illegitimate extrapolation of our own experience" (Rovelli 2018, 44).

But the idea of the present as something robust, perhaps the very bedrock of our experience of the world, is deeply ingrained in our mental attitudes. At the very least, it is thought, what I am experiencing now is real. This is what Whitehead called "presentational immediacy", "the familiar presentation of the contemporary world" (Whitehead 1930, 13), which he conceived in terms of a moment shared between a percipient and the things perceived. The relativity of the present undermines the idea of such a unique present moment—if that moment is assumed to be without duration, an instant. (This accounts for Whitehead's rejection of such an instantaneous now in which events occur, as discussed in Chap. 3.) It also appears to preclude the doctrine of presentism, according to which what exists is just what exists now, and equally, the common conception of becoming as consisting in the advance of such an instantaneous present from the past to the future, as in the "growing block" model we treated in Chap. 3. But this then raises the question: *what is the status of the present in relativity theory*?

One answer is to deny the present any objective significance, as in the static block model, where the present is relegated to being a mere feature of subjective experience. We have already discussed and rejected this model in Chap. 3, where it was found to commit an equivocation about the reality of events. But closely related to this view is what may be called the *relativized* or *observer-dependent present*, the first of three conceptions of the present offered in response to the relativity of simultaneity that I wish to consider in this chapter:

(i) *the present is relative to the observer's inertial frame.*

On this view, *present existence* is not an absolute notion, but *is relative to the observer's inertial frame*; the world-at-an-instant is a three dimensional, but relative, reality. This is a popular view, especially among physicists, who (taking their cue from Einstein) have concluded that the present is "relative to the observer's coordinate system".[4] Since they have also remained wedded to the idea that becoming

---

[3] Actually, we need to make a distinction between Newtonian and Galilean spacetime. According to Newton, each point x in absolute space marks the same position through time, making the resulting spacetime a unique structure, absolute space time. But according to the Galilean transformations, a boost in velocity v will make a state of affairs described in terms of the coordinate x indistinguishable from one described in terms of the transformed coordinate $x' = x - vt$. Since v is an arbitrary inertial velocity, this means there is an infinite class of equivalent spaces, each associated with a different value of v, in each of which x' marks the same position through time: this is Galilean spacetime.

[4] Thus Stephen Hawking writes: "In the theory of relativity there is no unique absolute time, but instead each individual has his own personal measure of time that depends on where he is and how

must be conceived in terms of the advance of a world-wide now, proponents of this thesis of the observer-dependence of the present also conceive becoming itself as observer-dependent.

The observer-dependence of simultaneity and becoming was contested by Alfred A. Robb as early as 1911 in a direct reaction to the way Einstein described his theory. Basing his theory on the Principle of Retarded Action, according to which influences must always take a finite time to propagate (see Chap. 4), Robb argued that no distant event could be simultaneous with any given point-event. He therefore insisted that:

(ii) *what is present at a given spacetime point is (strictly speaking) constituted by that point alone.*

This is the *punctual present* view, later taken up by Milič Čapek, and carefully elaborated by Howard Stein in some influential papers. Although formally correct, this conception of the present does not sit well with our experience of objects around us as present to us. Following certain suggestions made by Stein, Steven Savitt and I have independently suggested that a more natural notion of the present in relativity, and one appropriate to an event of finite duration, endows it with a certain spatiotemporal extent. The region of spacetime in question was first defined by the Russian physicist A. D. Alexandrov, so we may call the corresponding notion of the present the *Alexandrov present*:

(iii) *present to an object or process during an interval of its proper time π, are all those processes in the region of spacetime comprised within the absolute future of the beginning of this (usually short) interval and the absolute past of the end of that interval.*

My discussion proceeds as follows. I begin by stressing the need for the distinction between what is in an observer's (visual) experience and what such an observer infers to be 'now'. In fact, even classically, what you see—the events appearing in phenomenal experience—are all events that occurred in the past. This already creates difficulties for the kind of primacy accorded to the events of present experience. So before examining the status of the present in the light of relativity theory, it behooves us to take a step back, and examine pre-relativistic conceptions of the present, the topic of Sect. 6.2. We begin by examining Einstein's own contrast between time in relativity and the classical conception. Contrary to what seems to be implied by his account, the distinction between what *is* happening now and what *is seen as* happening now is not a discovery of relativistic physics, but is already taken into account in classical physics with the Doppler effect.

This occurs only in Einstein's attempt to situate the theory philosophically, and does not affect the substance of his theory. Nevertheless, it is intimately linked with his "interpretation of coordinates and time as products of measurement" through which they are connected with "immediate experience" (Einstein 1954, 247). A similar interpretation in his 1905 paper provoked various misconceived attempts to refute

---

he is moving" (Hawking 1988, 33). "Thus time became a more personal concept, relative to the observer who measured it" (143).

his theory on the part of his contemporaries. For according to this epistemological reading, local time is given its meaning, even if mediately, by its relation to the possible experiences of a putative observer. In Sect. 6.3, I show how analogous mistaken conceptions of relativity as "relativity to the observer" underlie the idea of the observer-dependent present. But one of the main lessons of relativity is that there is no instantaneous present "seen" by any observer. As in classical physics, an event *seen* as happening now by an observer in relative motion to it differs from what the same observer would *infer* to be happening now if they were relatively at rest; but in relativistic physics this effect is compounded by the effect of time dilation. I use a formulation of the scenario of the Twin Paradox to show the contrast between what the twins would see and what they would infer classically (because of the Doppler effect) and relativistically (because of the combination of that with the time dilation effect). As we shall see, in relativity distant events simply are not instantaneously *present* to an observer at a spacetime point in any meaningful sense.

A realization of the latter fact underlay the response of Alfred Robb to the problem of the present, which he offered in his (1911) and subsequent publications. In Sect. 6.4 I argue that Robb's view is of particular interest because he shows how Minkowski spacetime can be constructed from an invariant relation of chronological precedence, without the measuring rods and clocks assumed by Einstein. As clarified by later authors, Robb's construction can be seen as built around a principle of local becoming from which, together with the principle of relativity and certain natural constraints on physical process, the Galilean and Lorentz transformations can both be derived (as I sketch in the Appendix to this chapter).

In ascribing fundamental reality to isolated point-events, however, Robb's position is vulnerable to the objections raised in Chaps. 2 and 3 to Grünbaum's attempt to constitute reality from point-events. Robb's construal of the present is further vitiated by his insistence that his instants (point-events) be immediately accessible to consciousness. As I argued earlier, a point-event is better regarded as the bound of a process or event that has some duration. In keeping with this, finally, I argue in Sect. 6.5 that it is more appropriate to characterize the present in terms of what is compresent to such an extended event or process. This gives a view more consistent with our experience of compresence by allowing for an extended time of apperception. It yields an account of the present in the sense that you might say you were present at some significant event; but it is not intended to capture or reconstitute the present of phenomenal experience.

6.2 Pre-relativistic Conceptions of the Present

In a semi-popular essay he wrote in 1936 Einstein gave an account of how time in relativity theory differs from its construal in classical mechanics. He proposed that there are two independent postulates constituting the idea of the objective time presupposed in classical physics. One correlates "the temporal sequence of experiences with the readings of a 'clock', i.e., of a periodically recurring closed system", and this constitutes the core meaning of what he calls “local time”.[5] The second extends this notion to all the events in space, “by which notion alone the idea of local time is extended to the idea of time in physics” (Einstein 1954, 298). The first postulate envisages an observer with a clock, the material embodiment of time; the clock is a measuring device giving local time its “empirical content”, a conceptual precedence of measuring instrument to concept that “corresponds exactly to the precedence of the rigid (or quasi-rigid body in the interpretation of the concept of space” (298–99). Prior to the advent of the theory of relativity, Einstein continued, the prevailing illusion was that “the meaning of simultaneity in relation to spatially distant events, and consequently, the meaning of physical time, is a priori clear”, and that “this illusion had its origin in the fact that in our everyday experience we neglect the time of propagation of light” (299). As a result, “we are accustomed … to fail to differentiate between ‘simultaneously seen’ and ‘simultaneously happening’”, with the consequence that “the difference between time and local time is blurred” (299).

Local time, then, is for Einstein the time of a given observer, established by the linkage between his experiences and the clock he has immediate access to, and then extended across the whole of space by an appropriate measurement procedure. Where classical mechanics erred, according to him, was in supposing that extending any one observer’s local time across the whole of space in this way would yield a unique class of simultaneous events, whatever the observer’s state of motion—just as to untutored common sense there appears a whole class of simultaneous events that are immediately disclosed to sight. As a result of this blurring of the difference between time and local time, Einstein charges, the “lack of definiteness in the concept of objective time” was “veiled” or hidden” in classical mechanics, where time was represented as “given independently of our sense experiences” (299). Einstein’s account implies that objective time should be grounded in the observer’s sense experiences by a proper extension of his local time, the clock in his own reference frame, across the whole of space. Once this is done appropriately, it reveals that objective time lacks definiteness, because of its dependence on the observer’s own state of motion. Simultaneity is determined by the observer’s sending out and receiving reflected light signals, and the results will vary with the observer’s own state of motion. Each observer has his own local time. There is no absolute simultaneity.

Here Einstein is drawing a picture of the classical theory of the present by way of contrast with time in relativity theory, and at the same time giving an epistemological account of why the latter theory is superior. The Lorentzian view presupposes that there is one absolute time in which all events occur, and this is plausibly the result

---

[5] Albert Einstein, “Physics and Reality” [1936]; from *Ideas and Opinions*, (Einstein 1954, 298). Confusingly, this is not the same concept that Lorentz and others called *local time*, which is what results from the extension of local time across all of space, as in Einstein’s second postulate here. Of course, we are also familiar with ‘local time’ in another sense, as remarked by Alfred Robb. This is the sense in which “noon at Greenwich and noon at New York are both described as 12 o’clock local time, although the instants referred to are clearly different” (Robb 1921, 7). But as Robb explains, the idea underlying ‘local time’ in this sense is that different names are used for the same time at different locations on the Earth’s surface, just for the convenience of keeping clocks roughly synchronized with the daylight hours in those locations.

of classical prejudices resulting from our regarding nearby events as simultaneous because they are seen to be so, owing to the very great speed of light. In fact, however, as Einstein would have known very well from his work in the Patent Office, the speed of transmission of light signals must be taken into account in ascertaining the times of distant events. Nevertheless, the way Einstein has set up this contrast with classical physics implies that the classical foundations are faulty in not having taken this into account in determining the simultaneity of distant events, thereby failing "to differentiate between 'simultaneously seen' and 'simultaneously happening'". As we shall see, this is far from the case, and we will need to get this straight to appreciate the classical theory of the present.

Secondly, there is the epistemological grounding that Einstein provides. This 1936 article reflects Einstein's thoughts at a comparatively late period in the development of his thought, some twenty years after he had created his General Theory of Relativity, and over thirty after he had introduced the Special Theory in 1905. Yet the notion of local time he is advocating here, and especially the idea of its receiving its meaning from how it is measured, reveals a continuing commitment to the epistemological considerations that led him to the Special Theory. This is curious given that long before 1936 Einstein had repudiated such Machian considerations as a mere dispensable scaffolding. But it is significant in showing him still wedded to the view that there is no time in relativity theory independent of a subject and his or her state of motion, only a time relative to the observer's coordinate system.[6] This, as we shall see, has resulted in some serious misunderstandings of time in relativity.

Already in 1911, Alfred Robb had objected to the idea that relativity makes simultaneity observer-dependent. Einstein's claim "that events could be simultaneous for one observer but not simultaneous for another moving with respect to the first", was interpreted by him as giving "an air of unreality to the external world which cannot be justified" (Robb 1911, 11). For Robb, Einstein's idea that the simultaneity of spatially distant events depends on the observer's point of view "seemed to destroy all sense of the reality of the external world and to leave the physical universe no better than a dream, or rather, a nightmare." (Robb 1921, v). What Einstein was later to characterize as the "lack of definiteness" of objective time was for Robb rather a sign of the failure of simultaneity to apply to spatially distant events at all.

Robb's "nightmare", it should be noted, is something of a caricature. For Einstein, the "subject" could in fact be any material body in motion. But the importance of Robb's rejection of the subjectivism he saw in Einstein's thinking is that it led him to provide an observer-independent foundation for relativity based on an invariant relation of temporal precedence, and a theory of the present distinct from the classical idea of a plane of simultaneity. Moreover, in the introduction to his 1911 paper Robb appeals to the very distinction that Einstein would later invoke (in 1936) between the time we see an event to occur and the time at which it occurs. So a comparison

---

[6] Rugh and Zinkernagel make a similar point: "But even if Einstein subscribed less to a positivistic philosophy over the years (Ryckman 2001), he kept referring to the role of measuring rods and clocks in order to give physical meaning to the coordinates in special relativity" (Rugh and Zinkernagel 2009, 7). (The article referred to is updated as (Ryckman 2014).) See Howard (2017) for an analysis of Einstein's philosophy of science.

of Robb's use of this distinction with Einstein's will serve as a convenient way to introduce discussion of the classical theory of the present, and how it should be contrasted with the situation in relativity.

In his paper Robb discusses a hammer blow seen from a distance, and how auditory and visual perceptions of it differ in time "when the occurrence is at a distance from the observer" (Robb 1911, 8). He cites Ole Rømer's experiments from 1675–76 on the eclipses of Jupiter's satellites which showed the speed of light not to be infinite. "Thus the instant of one's visual perception of a distant event cannot be identical with the instant at which the event occurs, and we perceive near and distant events simultaneously which certainly do not occur simultaneously." (8). Robb concludes that, given the finite velocity of light, and the fact that with our other senses we perceive distant events more slowly still, all the distant events we perceive are actually in the past. So, if what is happening now is what is happening at this instant that we are perceiving, then, strictly speaking, "*the present instant does not extend beyond here, and the only really simultaneous events are events which occur at the same place.*" (18).

As described so far, Robb's argument depends only on the finite velocity of light, and not on any features unique to relativity theory. But this means that the same argument could have been conducted classically: nothing that appears to our senses is happening now, save what is coincident with us at that point in space and time. So if what is real is what is now, then reality is restricted to that fragmentary vantage point. Robb himself does not go that far, rejecting the idea that reality does not extend beyond the observer's own immediate perceptions as solipsism.[7] But we do find a non-solipsistic version of such a view of reality advocated over two thousand years ago by the Buddhist sect of Sautrãntikas, mentioned in Chap. 2 in association with time atomism. As explained by the Buddhist scholar Fyodor Shcherbatskoy, they proposed that reality is constituted by fragmentary, instantaneous point-events or *dharmas*, a doctrine known as the "Theory of Instantaneous Being". This is "the leading idea of Buddhism—there is no other ultimate reality than separate, instantaneous bits of existence" (Shcherbatskoy 1932, 1, 80). Although the origins of this view are lost in the mists of time (and may, according to Shcherbatskoy, even predate Buddhism itself), it has taken different forms in the different sects making up the chequered history of Buddhism. Originally it was a metaphysical thesis, devised in opposition to the Realists, who held that space and time were as real as the matter contained in them. But later it was given an epistemological foundation by Dignâga and Dharmakîrti (Shcherbatskoy 1932, 1, 108–114). If the only events that are real are those that are present to our perceptions, then, given that these perceptions themselves take time, a little reflection reveals that every event we perceive as 'now' actually took place earlier, and so is not strictly speaking still in existence. From this they concluded that what is past, future, or imagined is all alike unreal, and only the

---

[7] "A normal individual who is not a solipsist (and a solipsist could hardly be regarded as a normal individual) believes in the existence of more than his own self and his own perceptions, and one is accustomed to regard these perceptions, under normal circumstances, as representing things as real as one's self but in some sense external." (Robb 1911, 7–8).

instantaneous, punctual present is real.[8] Admittedly, if perceptions are conceived as instantaneous, and light is thought to travel instantaneously, this would not follow: at least everything we see at a given instant could be understood as existing now. But there are indications in the Nyaya texts that the Buddhists, like the Epicureans, held that the transmission of light, although almost inconceivably rapid, is not in fact instantaneous.[9]

One thing this shows is that there is no such thing as a univocal "classical theory of the present" shared by all authors prior to Einstein.[10] But it shows more: it shows that the common tendency to apply the label "real" only to those events that we are perceiving now—the doctrine of presentism—is already deeply problematic in pre-relativistic philosophy. If everything that is real consists only in my present perception of it, as Berkeley assumed with his doctrine of *esse est percipi* ("to be is to be perceived"), then nothing beyond my present perception can be real, including the chairs or trees I see around me, or the self I take to be perceiving them.[11] Even Hume, whose empiricist principles led him to reduce the self to a succession of instantaneous impressions, presupposes a continuous time in which these impressions can form a succession—an idea which cannot be derived from sense impressions.[12] There are similar difficulties for Kant, for whom time is "the form of inner sense". For if there is no continuously enduring self possessing this form of inner sense, contrary to presentism, then there is nothing that could be continuously producing the appearances of things in time.[13]

---

[8] Shcherbatskoy summarizes Buddhist presentism as follows: "Everything past is unreal, everything future is unreal, everything imagined, absent, mental… is unreal. Ultimately, real is only the present moment of physical efficiency [i.e., causation]." (Shcherbatskoy 1932, 70–71).

[9] South Asian History, Pages from the history of the Indian sub-continent: Rational and Secular Philosophy, Logic and Epistemology, Philosophical Development from Upanishadic Metaphysics to Scientific Realism: "Nyaya texts on causality indicate that there was an awareness that light travelled at a very high speed but the transmission of light was not instantaneous." (http://india_resource.tripod.com/upanishad.html).

[10] Craig Callender, for example, thinks it obvious that "Traditional conceptions of time all accept the non-uniqueness assumption", which he characterizes as requiring that "at least one event in the universe shares it present with another event's present" (Callender 2000, S592). Although this is perhaps implicit in Aristotle's notion of τo νυν, whose identity across space was made explicit by Gassendi and Newton, the example of the Sautrãntikas shows that it cannot be held to be universally true of traditional conceptions.

[11] One might add: since (according to orthodox theology) God does not perceive events "now", Berkeley's stratagem of trying to save the reality of things whose *esse est percipi* by allowing God as a perceiver will not suffice. One is therefore obliged either to deprive God of his atemporality, as in modern process theology, or, like Newton, to allow that God perceives all events in time directly by a kind of proprioception, as if in his own Sensorium; either solution would have been equally heretical to the mind of Berkeley.

[12] This point is made with characteristic wit, irony and modesty by Jorge Luis Borges in his essay (in two versions) "A New Refutation of Time", (Borges 1964, 190–205), in which he combines Berkeley's idealism and Hume's theory of self with Leibniz's Principle of the Identity of Indiscernibles to refute the reality of time.

[13] As we shall see in Chap. 8, Julian Barbour, taking his lead from John S. Bell, argues precisely this: there is nothing continuously producing the appearances of things in time: rather, the appearance

Einstein himself, of course, has often been situated in the empiricist tradition. Logical positivists attributed his success in overthrowing classical conceptions of space and time to his having correctly defined them in relation to experimental procedures. Percy Bridgman, for example, drew inspiration from Einstein for his doctrine of operationalism, according to which the meaning of a physical concept is synonymous with the operations by which we measure it. According to him, the key to Einstein's creation of the Special Theory of Relativity was his insistence on a proper operational definition of the concept of simultaneity (Bridgman 1927). As Robert DiSalle has observed (DiSalle 2006, 110), there is ample support for this in Einstein's original 1905 relativity paper, where Einstein sets out by giving two "operational" criteria for simultaneity: one could define it "by substituting 'the position of the small hand of my watch' for 'time'" (Einstein [1905] 1923, 39). But this will only define time "exclusively for the place where the watch is located" (39), so that for distant events this needs supplementing by a coordinative definition, in terms of "the time at which a light signal from each event reaches the watch" (DiSalle 2006, 110).

Recent studies of Einstein's philosophy of science have downplayed the significance of such positivist elements in his thought, showing how they played little role in his reconceptualizing of space and time, as opposed to its philosophical justification. Notably, however, the two operational criteria Einstein gave in his 1905 paper are the same two criteria for giving meaning to time that Einstein outlines in his 1936 paper described at the beginning of this section. As we have seen, there he insists that "the meaning of simultaneity" of spatially distant events cannot be determined except in relation to how it is measured, so that each observer in a different state of motion would have his own local time. So even though he later repudiates Machism, and insists that the creative formation of concepts is free and largely mathematical, he still insists that the meaning of time is different for differently situated moving observers.

This is not without significance. As I shall argue, Einstein's way of conceiving the meaning of time in terms of observers' possible experiences leads to a conflation of proper time with coordinate time that is still prevalent among contemporary physicists and philosophers. It has influenced not only those who saw themselves as following his lead, but even those who have opposed aspects of his natural philosophy.

A good case in point is provided by Whitehead, who can be regarded as having constructed an experience-based epistemology similar to that described by Einstein, but then used it to drive a metaphysics of time opposed to that implicit in Einstein's theories. Like the mature Einstein, Whitehead was critical of positivistic empiricism. Despite maintaining that philosophy should begin with an analysis of the content of subjectivity, he insisted that it needed to be corrected by reference to what he called the "reformed subjectivist principle", so as to show how objectivity is related to

---

of the past is encoded in a "time capsule" at any point in spacetime, yet there is no continuous time threading through these points.

subjectivity.[14] But this conception was substantially moulded by his understanding of relativity theory, where he followed Einstein in holding that it is a consequence of the theory that observers would construct different "meanings" for space and time depending on their state of motion.[15] Like Einstein, however, Whitehead rejected a subjectivist interpretation of this, insisting that these "meanings" are associated with the bodies of the observers, not their minds (Whitehead 1925, 118). As a result there are therefore "multiple time-systems", different "external" spacetime relations specific to observers or objects in different states of motion.

So Whitehead tries to construct time and spacetime from the differing possible experiences of observers. But where Einstein had inferred from his operational definitions that absolute simultaneity is an illusion, Whitehead parted company. For him, simultaneity is a basic component of experience, prior to objective temporal relations of succession: "simultaneity is an ultimate factor in nature, immediate for sense-awareness" (1920, 56). So, faced with the demise of the world-wide instant implied by relativity, his response was "to distinguish simultaneity from instantaneousness" (56), and to deny the instantaneity of simultaneity. "There is no such thing as nature at an instant posited by sense-awareness. What sense-awareness delivers over for knowledge is nature through a period" (57). This, he agrees with Bergson, is *duration*, namely "the whole simultaneous occurrence of nature which is now for sense awareness" (53).[16] It is in such an extended duration that "the creative advance of nature" occurs. A primitive sense of the passage of time is delivered by feelings of causal efficacy, and at a higher level, we conceive time as an extended serial order of moments. The latter is the objective time of physics, an abstract concept, produced by the activity of the experiencing subject, but actualized by the passage of nature.

This is not the place for a detailed analysis of Whitehead's complicated metaphysics of time. As an attempt to save the absoluteness of simultaneity in the face of

---

[14] Hurley (1986, 99). Although Whitehead interprets the experiencing subject in terms of "actual occasions" rather than the other way around, the world for him, as for Kant, is nothing more than what is disclosed in sense experience.

[15] "But now it seems that the observed effectiveness of objects can only be explained by assuming that objects in a state of motion relatively to each other are utilising, for their endurance, meanings of space and time which are not identical from one object to another." (Whitehead 1925, 120). Ironically, Whitehead sees himself as opposing Einstein's operational definition of simultaneity based on light signals, where "the very meaning of simultaneity is made to depend on them" (119), by one in which "the meanings of our terms" are "physical facts expressible in terms of immediate perceptions" (56). See (Bain 1998) for further discussion. Bain claims that Whitehead's insistence is a *semantic* one: "the manner in which theories are formulated and the types of concepts they employ should make clear what one's ontological commitments are". But the same applies to Einstein: see e.g. Ryckman (2014).

[16] Cf. Bergson: "We claim that a single time and an extension independent of duration continue to exist in Einstein's theory considered in its pure state; they remain what they have always been for common sense" (Bergson 1965, 30); and "While the duration of the phenomenon is *relative* for the physicist, since it is reduced to a certain number of units of time and the units themselves are indifferent, this duration is an *absolute* for my consciousness, for it coincides with a certain degree of impatience which is rigorously determined." (Bergson 1944, 368).

relativity, it simply will not work.[17] But I believe that there is a moral to be drawn from its failure. This is that the strategy of trying to construct a physical theory out of terms whose meanings are given through their being related to an observer's immediate experience is mistaken in principle. It inverts the methodological practice of physics, which does not *construct* its concepts out of subjective or phenomenological givens, but *infers* what is the case independently of knowing subjects. For the present of classical physics is an *inferred* present. Knowing the speed of light and sound and the time it takes to perceive events of various kinds, we can determine which events are those that could possibly be experienced by someone in a given location in classical spacetime, and, knowing the motions of distant bodies we can calculate also those events that are happening at the same time. We can infer, for instance, the position of each of Jupiter's moons at the very time I observe them through my telescope (in slightly different positions) if we know enough about their positions and orbital velocities. We can, that is, infer where they are "now".

This is precisely the consideration on which Rømer's determination of the speed of light is based.[18] He was able to establish the time taken for Jupiter's moon Io to complete its orbit around that planet from a series of observations of its successive emergences from eclipse by Jupiter's shadow. But if light travelled at a finite speed, the times at which it was seen to emerge from eclipse would differ when the Earth was at different distances from Jupiter. So, knowing the time of its emergences on August 7 and August 14, 1676, the times for its emergences in November could be calculated. The observed times differed from those predicted, allowing Rømer to calculate the extra time taken by light to reach the Earth from Jupiter in November 1676 compared to August 1676 as about ten minutes, giving the value of c as 2.2 $\times$ 10^8 m/s, as noted in Chap. 5. Rømer's calculations were confirmed by Fontenelle and later improved upon by Bradley. As was later recognized, it is implicit in these considerations that the frequency of an oscillating body (such as the orbiting Io) will appear to speed up as it approaches you, and to slow down as it moves away. This is the famous Doppler Effect, explained by the Austrian physicist Christian Doppler in 1842.[19] This effect is already noticeable with sound, as the pitch of an approaching emergency vehicle's siren rises at it comes towards you, and then falls as it recedes into the distance.

So the distinction between the perceived present and the inferred present was, as Einstein certainly knew, perfectly well established in classical physics. Aristotle's τo νυν, an indivisible now constituting the boundary between past and future, was part of the Western classical heritage (as we saw in Chap. 2). That such a moment extends spatially across the whole world is clearly articulated by Pierre Gassendi in

---

[17] Whitehead was no armchair philosopher, however: in an effort to give substance to his metaphysics of space and time, he constructed a theory of gravity on that basis that was a serious alternative to Einstein's General Relativity for many decades.

[18] For sources and thorough discussion see (Rømer and Cohen 1940).

[19] Although the idea of the effect is implicit in the calculations of Rømer et al., Doppler gets the credit for first discussing the changes of frequency of waves whose source is in relative motion explicitly in his treatise (Doppler 1842). It was confirmed for sound waves three years later by C. H. D. Buys Ballot.

the mid-seventeenth century: “As any moment of Time is the same in all places, so any portion of Place is in all times.”[20] Gassendi’s notion of a world-wide moment was subsequently adopted by Newton, for whom “each and every indivisible moment of duration is *everywhere*” (Newton 1999, 941). Moreover, as we saw in Chap. 4, Newton’s “absolute true and mathematical time” is distinguished from its sensible measures by the “equation of time”—that is, by the correction due to the relative motions of the bodies involved, the need for which is evidenced empirically “by eclipses of the satellites of Jupiter” (Newton 1999, 410). Newton was under no illusion that light travelled instantaneously, and in fact was among the first to accept Rømer’s establishment of the finite speed of light.

Thus it is not the case that classical physics erred in extending the time of any one observer’s immediate experience across the whole universe. This universal or absolute time is an inferred time, and its distinction from the times of events as experienced by observers is firmly enshrined in the Doppler effect. This was well known prior to relativity and the discovery of time dilation. The distortion of time involved in time dilation is a distinct effect; but it too is an inferred effect, even though, like the Doppler effect, it will manifest itself empirically.

In sum, we have seen that if we keep in mind the distinction between what is perceived or experienced and what is inferred, there is no confusion in the notion of a world-wide present in classical physics. It is perfectly normal in physics to infer things that go beyond our experience, so long as supposing them helps to explain our experiences, and obtains confirmation through them. The problem with the theory of the present implicit in classical physics is not that it posited the existence of simultaneous events which could not be part of our experience at a given instant. It was, as Einstein himself demonstrated, that the supposition of a unique class of such events turned out to be incompatible with two other well-confirmed empirical results: the independence of the speed of light from the speed of its source, and the principle of relativity.

6.3 The Observer-Dependent Present

Physicists commonly take relativity to mean that all motion is relative to the observer. If you are moving relative to me, then I am moving relative to you with the same relative speed in the opposite direction. But then by the same token, if your length is foreshortened for me, mine is for you, your time is dilated for me, mine for you. On such a view motion is observer-dependent, and so are all the quantities dependent on it. Thus simultaneity is relative to the observer’s coordinate system, and the objectively given present of classical physics must be rejected, as Einstein charged. Early critics of Einstein picked up on this implied lack of objectivity. In

---

[20] “Ut quodlibet Temporis momentum idem est in ominbus locis; ita quaelibet Loci portio omnibus temporibus subest.” Pierre Gassendi, *Animadversiones in Decimum Librum Diogenis Laertii* (Lyon 1649, 632); my translation.

1927 W. D. MacMillan, a respected professor and author of a number of physics textbooks, refused to accept the theory. According to him, it "played havoc with the fundamental concepts of the entire human race". He continued,

> Time is no longer a public matter, the same for all, but is a private, personal matter. Your time and my time are different, just as your personality and my personality are different. Two events, one in New York and one in Chicago, are simultaneous in my time, but if you are moving with respect to me these two events will not be simultaneous in your time.[21]

More famously, the astronomer Herbert Dingle, at the end of a distinguished career in which he had served as Astronomer Royal and had written books explaining relativity theory, became an outspoken critic. If each of the twins in the twin paradox scenario is moving relative to the other, he reasoned, then each's life-process will be slowed down relative to the other, because of time dilation. But then each will age less than the other, an obvious impossibility![22] He was unfazed by experimental confirmation of Einstein's theory: "surely, one does not need an experiment to prove that one clock cannot at the same time work both faster and slower than another" (Dingle 1972, 19).[23]

But there is no impossibility here, as we saw in the previous chapter. The difference in the twins' ages when they reconvene is real, not observer-dependent, notwithstanding the symmetry between their situations when they are in relative inertial motion. Indeed, Einstein himself, despite his talk of the "lack of definiteness" of time in relativistic physics, nevertheless regarded time dilation as a physical effect, not just a "personal matter". As we saw, it was correctly predicted by him as a result of special relativity. Indeed, he suggested an experimental test of time dilation, concluding that "a balance-wheel clock that is located at the Earth's equator must run very slightly slower than an absolutely identical clock, subjected to otherwise identical conditions, that is located at one of the Earth's poles" (Einstein [1905] 1923, 904–905). The same goes for Langevin. In fact, it was precisely to try to forestall the kinds of objections

---

[21] W. D. MacMillan, "Postulates of Normal Intuition", in (Carmichael 1927, 54). MacMillan (1871–1948) wrote standard textbooks such as *The theory of the potential: Theoretical mechanics* (1958)*, Statics and the dynamics of a particle*, *Dynamics of rigid bodies: Theoretical mechanics* (1960), and *Periodic orbits about an oblate spheroid*.

[22] Dingle (1890–1978) made this objection in his introduction to the translation of Bergson's *Durée et simultaneité* (Dingle 1965, xvi). This was part of a quixotic campaign to refute relativity, culminating in his *Science at the Crossroads* (1972). But although these works appeared a whole generation after MacMillan, Dingle had already made his essential objection to time dilation in *Nature* (Dingle 1939) while still accepting relativity: "In an article in *Nature* I claimed that the twins must necessarily age at the same rate because it was an essential requirement of the special theory of relativity, which I then believed to be sound, that no observation was possible that would enable one to ascribe the motion preferentially to either twin." (Dingle 1972). See Marder (1971) for a thorough debunking of such misunderstandings about relativity.

[23] Similar criticisms were made by Bergson (1965, 74–78) and Lovejoy (1931). In his discussion of them Čapek (1971, 245–256) endorses their wholly mistaken view that "the time-retarding journey is meaningless within the special theory" (246), having applauded Dingle's claim that it is a contradiction for each of two clocks to be time-dilated with respect to the other (241).

made by MacMillan and Dingle that he devised the “Twin Paradox” scenario (and resolved it) in 1911, as we already noted above.[24]

Nevertheless, the idea that the present moment—and thus what is real—is relative to the observer is still prevalent. If the present is observer-dependent, so it is argued, and what is real in an observer’s experience is what occurs in the present instant, then what is real can vary with the observer’s state of motion.[25] We noted examples of this line of thought in Chap. 5. Thus Hilary Putnam begins his 1967 article with the principle that “There Are No Privileged Observers”, defined as follows:

> If it is the case that all and only the things that stand in a certain relation *R* to me-now are real, and you-now are also real, then it is also the case that all and only the things that stand in the relation *R* to you-now are real. (Putnam 1967, 241)

In the context of Special Relativity, he claims, “we have to take *R* to be the relation of simultaneity-in-the-observer’s-coordinate-system”.[26] Combining this with the principle that “All things that exist now are real”, he is able to show (supposing sufficiently many equivalent “observers” at distinct spacetime locations) that all events are equally real. If being real is synonymous with being present to an observer, and there are no privileged observers in the sense just defined, then “the notion of being ‘real’ turns out to be coextensive with the *tenseless* notion of existence”, where the latter “amounts to ‘will exist, or has existed, or exists right now’” (247).

Evidently something has gone wrong here. For if any event that exists now is real, and this “amounts to” its existing before now, existing now, or existing later than now, then an event’s existing now (say, at a specific time coincident with my utterance) is not distinguished from its existing earlier or later than my utterance; but that is tantamount to denying temporal succession.[27] In depicting reality as a relative to the observer, Putnam has made the same move as MacMillan and Dingle. If all observers are on a par, Dingle argued, then each observer has the same right to say that the other’s time is dilated. But since the same clock cannot go faster and slower than another, time dilation cannot be real. An analogous reasoning underpins Putnam’s argument: if what is real is what is simultaneous to the observer, then what is real for one observer will be past or future for another. So, since all observers are

---

[24] The Twin Paradox is often called the “Clock Paradox”, and ascribed to Einstein. But as Galina Weinstein has argued, this involves a confusion. What Einstein described and predicted in his 1905 paper is *time dilation*, the retardation of a clock in relative motion. Certainly the very notion of time dilation was regarded as paradoxical by those like MacMillan and Dingle, given the relativity of motion. But it was Langevin who devised the *Twin Paradox* scenario in an effort to forestall such misconceptions. See Weinstein (2012).

[25] Cf. Brian Greene: “*Observers moving relative to each other have different conceptions of what exists at a given moment, and hence they have different conceptions of reality*” (Greene 2004, 133–34).

[26] Here I have silently corrected a mistaken duplication of ‘the’s in the original: “we have to take *R* to be the the relation…” (Putnam 1967, 242).

[27] This conclusion is somewhat disguised by Putnam’s interpreting his position as supporting “the tenseless notion of existence”; for the latter is not supposed to be opposed to distinct events *existing* at different times, to temporal succession, but just to their *coming into existence* at those times. See Chap. 3.

on a par, all these events will also be real. But Putnam should have followed this logic to its conclusion, and inferred that, if the same event can be both in the future (or past) and simultaneous, then temporal succession cannot be real.

The mistake in each case is the same. It is to interpret relativity to an inertial system of reference as being synonymous with relativity to an observer, and then imposing an equivalence of observers' experiences. If two bodies a and b are moving uniformly relative to on another, then all processes occurring on b will be time-dilated relative to a, and all processes occurring on a will be time-dilated relative to b, whether or not there is anyone to observe the dilation. So long as there is relative motion between them, there will be relative time dilation of the processes in one with respect to the other—and this will be so even if we adopt any other inertial frame from which to observe the motions. Although it is true that if we adopt as our inertial frame that in which a given observer is at rest, we will get a picture of how things will appear to that observer, *there is nothing sacrosanct about the observer's rest frame*. Any one observer's frame (any inertial frame) will do just as well to describe the physics, but we should not switch frames while describing the same process.[28]

Just such a switching of frames is appealed to in many accounts of the twin paradox, under the mistaken conception that everything must always be described from a point of view in which the observer is at rest. Such a conception goes along with the idea of a reference frame as defining what is present in the observer's experience. On this reading, what is real in an observer's experience will be all those events simultaneous with the observer at each moment. In actual fact, however, the point-events that are simultaneous with the observer at any instant are precisely *not* events the observer could experience at that time! Here again there is a confusion of the question of what events may be *inferred* to be simultaneous at an instant, with the question of which events *are experienced* at that instant.

We can clarify this by re-examining the twin paradox under the adoption of an Einsteinian cosmos, first on the assumption of an observer-dependent present, and then by carefully distinguishing what the twins *see* from what they *infer*. Accordingly, we now assume that the length of an object moving at speed v relative to an observer is foreshortened in the direction of motion by the factor $\gamma = 1/\surd(1 - v^2/c^2)$, and times are dilated by the factor $1/\gamma = \surd(1 - v^2/c^2)$. The mathematics is pretty straightforward. We suppose one twin is Terence, who stays on Earth for twenty years, while his astronaut sister Astrid travels to Alpha Centauri (taken as 6 light years away) at a speed of $0.6c$, and back again to Earth at the same speed. Ignoring any periods of acceleration or deceleration, according to Terence, his twin Astrid is away for exactly twenty years, ten years out, ten years back. Things are otherwise

---

[28] Cf. Howard Stein: "one often characterizes such notions as simultaneity, time interval, and length, in Einstein's theory, as 'relative to the observer'. As we shall see, this way of speaking can lead to serious confusion. One should not lose sight of the fact that any observer, regardless of his state of motion, can express the results of his own measurements (and those of colleagues) in *absolute, physically meaningful language*, which is *the same for all observers*." (Stein 1968, 12); cf. also "an observer's state of motion does not impose upon him, according to relativity theory, a special view of the world's structure. The view that it does—that the observer has some distinguished subjective or perceptual relation to "his present"—leads only to confusion." (Stein 1970, 293–4).

for Astrid. At such great speed, the distance between the Earth and Alpha Centauri is for her foreshortened by a factor $\sqrt{(1 - 0.36)} = \sqrt{(0.64)} = 0.8$. So she needs to make a journey outwards of only 4.8 light years. An easy calculation shows that, travelling at a speed of $0.6c$, she does it in $4.8/0.6 = 8$ years according to her clock. The distance home is equally foreshortened, so that she takes only 16 years for her journey (still imagining zero turn-around time). Thus when they are reunited and compare their clocks, they find that Terence has aged 4 years more than his sister!

You can see why Dingle found this paradoxical. If all inertial motion is relative, how can there be an absolute difference in their lifetimes resulting from it? While the twins are in relative inertial motion, each's duration will be running slower from the perspective of the other's rest frame. In each leg of the journey, Astrid would infer processes to be happening more slowly on Earth as it receded from her or approached her at $0.6c$: her eight years would correspond to an inferred duration of processes on earth of only 6.4 years!

There is, of course, an asymmetry in the scenario. In order to reunite with her brother, Astrid must turn around, and this requires that she should accelerate. So something must happen at the point at which she switches direction that explains the difference in their lifetimes. For so long as the twins are in inertial motion relative to one another, each twin must indeed infer that the other's clock is running slow. Thus it is often said that the reason for the discrepancy in the twins' ages is that whereas Terence is in inertial motion throughout, Astrid is the one who really moves because of her acceleration, although this acceleration lies outside the scope of the theory. This has led many to claim that a proper resolution of the paradox must appeal to General Relativity, since Special Relativity (SR) applies only to systems in inertial motion.[29] There are indeed statements by Einstein to this effect, and these are probably the ultimate source of numerous similar such claims that can readily be found.[30] But we know this to be incorrect. Special Relativity is perfectly applicable to accelerated motions, which are represented as curved worldlines in Minkowski spacetime. Granted, in the toy model we are considering, the path of Astrid is not so much curved, as kinked. An instantaneous turnaround is unphysical, but this does not alter the fact that we can model it in Minkowski spacetime; also, one can make the same case with a curve instead of a kink.

---

[29] For example, Max Born held this to be the case in his (1922, 216). Similarly, J. S. Bell writes that "accelerated observers are not considered in the 'special' theory" (Bell 1987, 77), and Milič Čapek that "It is this acceleration which places the '*voyage au boulet*' clearly beyond the purview of the special theory. Thus only within the framework of the general theory can Langevin's spectacular space trip be correctly analyzed." (Čapek 1971, 246). Čapek's conflation of proper time with time in an observer's rest frame vitiates his attempt to make sense of aspects of Bergson's failed attempt to come to terms with the special theory.

[30] Thus Brian Greene (in the course of an otherwise excellent exposition) writes that "in special relativity Einstein's main focus was on a special kind of motion: constant-velocity motion. It was not until 1915, some ten years later, that he fully came to grips with more general, accelerated motion, through his general theory of relativity." (Greene 2004, 51). One still finds similar statements on the Internet, e.g.: "However, this resulted in a limitation inherent in Special Relativity that it could only apply when reference frames were inertial in nature, (meaning when no acceleration was present)." (http://en.wikipedia.org/wiki/Inertia; accessed January 30, 2018).

But let us return to the idea of the observer-dependent present. If we assume that what is present (and therefore real) for Astrid is what is simultaneous with her in "her own frame of reference", then there must be some dramatic shift occurring at the point of her switching direction. What is real for her at the instant before she turns around at Alpha Centauri—all the events simultaneous with her at that instant—is radically different from what is real for her the instant she begins the return journey. We can calculate how it changes. The instant before she turns around, at the very end of her 8 years of travel, Terence's clock will read $8 \times 0.8 = 6.4$ years. Similarly, during her journey home Terence's clock will advance only 6.4 years, yet it will read 20 when she returns to Earth. So, at the instant she sets out on her journey home, his clock must read $20 - 6.4 = 13.6$ years. Assuming an instantaneous turn-around, events on Earth that are real for her would have had to have jumped 7.2 (=13.6–6.4) years from one instant to the next.

So on this account, the explanation for the discrepancy in the twin's ages is due to the fact that the acceleration discontinuously skews Astrid's temporal orientation. If we were to follow Putnam's informal way of speaking, we would say that Astrid "experiences" 7.2 years on Earth going by in an instant: events that were "present according to her co-ordinate system" are discontinuously displaced 7.2 years into the past of "her-now" according to that same system.[31] In actual fact, however, no such wrenching change of her experience of any events would occur. She does not *see* any of the distant events that are simultaneous with her, either before or after her turn-around. There is nothing "present according to her co-ordinate system". These are facets of a sloppy use of the ideas of "observer's reference frame" and the observer's "present", and a failure to distinguish between the time an observer might *infer* an event to occur from when the observer would *see* it occurring.[32]

In fact, it will be worth going over the whole thing in some detail to see how this could be the case, but this time carefully distinguishing what an observer sees from what that observer infers.

First, let's examine the case of the interstellar twins from the point of view of classical physics. We assume, as did Rømer, that light travels (from the standpoint of Earth) at a finite speed, and that everything takes place in Newton's absolute time. Astrid will take 10 years to arrive at Alpha Centauri as before. But because light takes 6 years to travel from Alpha Centauri to Earth, when Terence actually sees the event of Astrid's arrival there, 16 years will have passed since they parted! He sees Astrid's clock register only 10 years while his has registered 16. Thus Astrid's clock is will appear to be running slow by a factor 5/8 compared to his, as he will have been noticing throughout her trip, with his telescope trained on her clock. This is the

---

[31] Cf. also Rietdijk's talk of two observers mutually at rest "experiencing the same 'present'" as one another. As Howard Stein has remarked, "there is of course no such 'experience': the fact that there is no experience of the presentness of remote events was one of Einstein's basic starting points" (Stein 1968, 16, n. 15).

[32] Cf. Sklar (1974, 272): "One must always be careful in special relativity to distinguish what an observer actually sees, literally, from what he computes to be the case"—the sloppiness of this slippage from one to the other perhaps constituting what one of my students called "the fallacy of the slippery slop"!

classical Doppler effect, as noted above. Now, since his sister takes the same 10 years to make her return trip, she is back after 20 years, and thus after only 4 more years by his clock. But when they reunite, their clocks will agree that 20 years have past. So during this 4 years he watches his sister's clock running at 10/4 = 2.5 times the speed of his own. Thus as Terence views Astrid's return trip and all the processes happening in it, he sees them appearing to occur four times as fast (2.5 divided by 5/8) as during the trip outwards!

Astrid has an analogous experience. When she arrives at Alpha Centauri, she observes Terence's clock to be reading only 4 years. For the image of Terence's clock registering 4 years travels the 6 light-years to Alpha Centauri to arrive there 10 years later (since light is travelling at speed c in Terence's frame of reference). So Astrid sees Terence's clock has been running 2.5 times as slowly as hers (i.e. at 2/5 speed)! But in the ten years it takes her to return home, Terence ages $20-4=16$ years, and accordingly on the way home she sees his clock to be running 1.6 times as fast! Thus she, too, is puzzled to see her twin's clock going 4 times as fast on her journey home (1.6 divided by 2/5) as it was on the outgoing leg of the journey.

All this is in accord with the classical Doppler effect. Events and processes occurring in a frame of reference in motion towards an observer appear to be speeded up ("blue shift"); occurring in a frame of reference in motion away from an observer, they appear to be slowed down ("red shift"). None of this has anything to do with time dilation. Everything here is purely classical, and the slowing down and speeding up of clocks observed by the each twin is an appearance only, completely explained by the Doppler Effect. Each twin, moreover, is able to infer what is happening "now" at any given time, and to distinguish it from what he or she is observing as happening now at that time. For instance, Astrid knows that at the moment she reaches Alpha Centauri, Terence's clock will be recording 10 years as having passed, even though she sees it recording only 4. The distinction between what is happening now and what is seen as happening now may be "blurred" in our unreflective everyday experience, but it was not blurred in classical physics, as one might take Einstein's discussion to have implied (Einstein 1954, 299).

Now, not everything about the twin's experiences is symmetrical. Although each twin sees the other's clock running 4 times as fast on the return trip as it was on the way to Alpha Centauri, they did not observe the same degree of red shift while moving apart, and the same degree of blue shift when moving back towards one another—Terence saw a 5/8 red shift while moving apart and a 5/2 blue shift while they were approaching, whereas Astrid saw a 2/5 red shift and an 8/5 blue shift. That they didn't observe the same shifts is explained by the fact that we have assumed that the light is travelling with the speed c in Terence's frame, not Astrid's. But if all inertial motion is relative, they should have experienced exactly the same degree of red shift while moving apart, and the same degree of blue shift when moving back towards one another. So this gives a nice illustration of the predicament facing physicists at the beginning of the twentieth century. All the experimental evidence seemed to suggest that the speed of light is independent of the speed of the source. But this is incompatible with the requirement that all inertial frames are on a par with respect to the laws of physics, unless something else gives. As we saw in Chap. 5, various physicists realized that appearances would be preserved if the length of a body were foreshortened in the direction of its motion through the aether by a factor of $\sqrt{}(1 - v^2/c^2)$. We have already seen how that resolves the situation about what they infer. But what do they see?

When Astrid arrives at Alpha Centauri, Terence's clock registers 16 years—10 years for her to make the trip, and 6 years for the image of her arrival to reach his telescope on Earth. But, because of time dilation, when Astrid arrives her clock actually registers only 8 years, so the image of the clock hand at 8 years is what Terence sees. So her clock *appears* to Terence to be running slow by a factor of 2, i.e. to be going half as fast as his; and it would have been appearing to go at half speed throughout this leg of the journey. On the return leg, Terence sees his sister's clock advance 8 years whilst his only advances the remaining 4 before their reunion; so the clock (and the aging processes of his sister and everything moving with her) *appear* to be running fast by a factor of 2. Despite these appearances, of course, Terence can *infer* that in each case the effect is 0.8 times what would be expected from the Doppler effect alone: a lag by a factor of 5/8, multiplied by 0.8, gives 1/2; a speeding up by a factor of 2.5 times 0.8, yields 2. Thus he *infers* that Astrid's clock is running slow because of time dilation.

Astrid, on the other hand, on looking back to Earth as she is arriving at Alpha Centauri 8 years later, sees the Tellurian clock register 4 years, as before. By her reckoning, the image has travelled eight years to get to Alpha Centauri, so Terence's clock *appears* to Astrid to be running slow by a factor of 2. On the return leg, Astrid sees her brother's clock advance its remaining 16 years whilst hers only advances 8; so the clock (and the aging processes of her brother and everything moving with him) appear to be running fast by a factor of 2. Again, she can calculate that since the effects should have been 5/8 and 2.5 if they were due to the Doppler effect alone, the difference is due to the fact that Terence's time is slowed relative to hers by the time dilation factor 0.8. (During her 8 years on the outward leg, she sees Terence's clock move 4 years when it should have moved 5 by the Doppler effect alone, since 8 times the Doppler effect of 5/8 = 5; on the way back she sees Terence's clock move 16 years instead of the 20 that would be 8 times the Doppler effect of 2.5).

Thus the situation is entirely symmetrical: while they are in relative motion, each twin suffers (relative to the other) an *inferred* time dilation, a slowing-down of the aging process. Moreover, each twin suffers precisely the same time dilation relative to the other while they are moving inertially relative to one another. The time dilation is a *real* but *relative* effect. And in terms of what *appears*, both see their twin sibling's clock running slow by a factor of 2 while they are moving apart, and running fast by a factor of 2 when they are approaching one another.

This shows us that the scenario depicted is entirely consistent. But how does it resolve the paradox? If everything is symmetrical, then why don't the twins age by the same amount? The correct explanation, as we saw in Chap. 5, is that it is not any difference between inertial frame-times that accounts for the difference in the twins' ages, but the difference in their paths through spacetime, the measure of which is proper time. It is the proper time elapsing along a particular path in spacetime that measures how fast the processes traversing that path are going, how fast the people or things undergoing them are aging, how fast they are becoming—and *it is this* *proper time that is the time told by the clocks throughout the twins' journeys*. In the non-Euclidean metric of Minkowski spacetime, it is the longest, not the shortest, interval between two spacetime points that is given by the straight line in spacetime connecting them—contrary to what a diagram of their motions might suggest. Astrid travels along two sides of a spacetime triangle, and Terence by the remaining side. In each case Astrid's path appears longer when we draw it, but the time elapsed (the integral of the proper time along the path) is shorter. It follows that it can't be said, as one often reads, that the duration of processes in SR is relative to an inertial frame. In the sense of time lapse that is relevant to the twin paradox—how much time elapses for each twin—*it is simply false that time lapse is frame-dependent*, i.e. that it depends on the inertial frame adopted. Indeed, the duration of each twin's journey through spacetime is an invariant measure: the proper time is numerically the same in all inertial reference frames, provided we use the same frame of reference to describe the whole journey.

Thus time does not lapse differently for the twins because of any difference they see in what events are present at any given instant—precisely because they cannot see events that are simultaneous with them at an instant! Whitehead, as we saw, followed Einstein in claiming that different observers would ascribe different "meanings" to time on the basis of their possible experiences. But as David Mermin notes, "That no inherent meaning can be assigned to the simultaneity of distant events is the single most important lesson to be learned from relativity." (Mermin 2005, xii). Similarly, Howard Stein: "the fact that there is no experience of the presentness of remote events was one of Einstein's basic starting points." (Stein 1968, 16, n.15). Distant events simply are not *present* to an observer at a spacetime point in any meaningful sense.

6.4 Robb and the Punctual Present

This brings us to the third of the construals of the present in relativity theory outlined above: what is present at a given spacetime point is (strictly speaking) constituted by that point alone. As we have seen, this view was first articulated by Alfred A. Robb in 1911, six years after Einstein's original 1905 paper, and only three years after Minkowski's.[33] It has since been endorsed by Čapek (1966), (1975), and also by Stein (1968) (although here without attribution to Robb), in their critiques of the arguments of Putnam and Rietdijk discussed in Chap. 5.[34] As we have seen, Robb took exception to Einstein's proposal that "events could be simultaneous for one

---

[33] Robb (1914); his (1936) is essentially a second edition of this book. Robb had previously published a draft of part of his theory in the short tract *Optical Geometry of Motion* (1911), and later gave a simpler exposition of his (1914) without proofs of the theorems in his (1921).

[34] "Like Rietdijk," objects Čapek, "Putnam retains the old notion of the universal present spread as a 'world-wide instant' across the whole universe, and uses this notion in order to conclude that, in a sense, *everything* is present" (1975, 612–13). But this neglects "the one essential idea of relativity that … 'Here-Now' can never be extrapolated to 'Everywhere-Now'" (613). Similarly, Stein objected that "if 'presentness to each other' of events is taken to mean that *each has for the*

observer but not simultaneous for another moving with respect to the first" (Robb 1936, 11; cf. 1921, v). Consequently, he regarded instants as restricted to spacetime points, and "avoided any attempt to identify instants of time at different places."[35] His instants are point-instants, and his events are point-events.

Robb presents his work as deriving directly from that of Larmor and Lorentz rather than Einstein.[36] This is a little eccentric. He cedes priority to Einstein and Minkowski for the mathematics, but objects to the subjectivism he sees as implicit in Einstein's way of presenting relativity in terms of observers, rods and clocks. For in his own construction of relativity theory Robb makes no use of rods or clocks at all.[37] Neither did Minkowski, of course. But whereas Minkowski's formulation of relativistic spacetime was premised on Lorentz invariance under Einstein's interpretation,[38] Robb derives the whole theory—spacetime and Lorentz transformations—from first principles. Moreover, Minkowski took Einstein's proposal that the local times corresponding to different inertial systems are ontologically equivalent to mean that time, like space, is a mere projection onto spacetime. Time is accordingly stripped of its former pre-eminence, as Minkowski proudly announced: "space by itself and time by itself recede completely, to become mere shadows". In Robb's system, by contrast, time takes pride of place: "the theory of space becomes absorbed in the theory of time, spacial relations being regarded as the manifestation of the fact that the elements of time form a system in conical order: a conception which may be analyzed in terms of the relations of after and before." (Robb 1914, 9)

Let me begin with a brief outline of Robb's theory and its significance for the reality of becoming in SR, before proceeding to his theory of the present.

In his 1914 theory, his "chronogeometry",[39] Robb takes as his starting point relations that for Minkowski are features of the "absolute world", consequences of the structure of spacetime. These are what Robb calls the "absolute relations" of earlier and later, described by Minkowski in the words: "Each worldpoint within the past lightcone of O is necessarily always earlier than O, each worldpoint within the future lightcone is necessarily always later than O."[40] Now what is in the future lightcone represents all the worldpoints that can be reached from O, and what is in the

---

other already become, then … in Einstein-Minkowski space-time *an event's present is constituted by itself alone*." (Stein 1968, 15).

[35] This is Robb's description of his 1911 tract in (Robb 1914, 3). In that work, as we shall see, Robb even wrote of one's being "directly conscious" of instants (1914, 8).

[36] "Although generally associated with the names of Einstein and Minkowski, the really essential physical considerations underlying the theories are due to Larmor and Lorentz." (Robb 1914, 1)

[37] "It is quite unnecessary, in our theoretical investigations, to introduce clocks or measuring rods at all." (Robb, unattributed; quoted from Winnie 1977, 134).

[38] As Minkowski writes, "The credit of first recognizing clearly that the time of one electron is just as good as that of the other, that is to say, that t and t' are to be treated identically, belongs to A. Einstein." (Minkowski [1908] 1952, 82; cf. 2012, 117).

[39] The term 'chronogeometry' is apparently due to A. D. Fokker, who happened on substantially the same insights as Robb without apparently knowing of his work. See (Fokker 1953/55), and (Alexandrov 1967, 1119).

[40] Minkowski ([1908] 2012, 119). Minkowski's calling them "necessarily always" earlier and later is doubly unfortunate: as we saw in Chap. 3 the relations could be said to hold "always" only if there

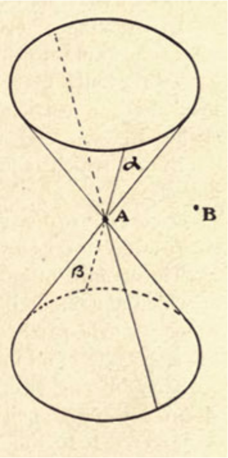


Fig. 6.1 B in the 'elsewhere' of A (from Robb (1914, 5)

past lightcone all those worldpoints from which any process could reach *O*—since no physical influence can propagate faster than light. Robb now takes the possibility of a physical influence going from *A* to *B* as a necessary and sufficient condition for *A*'s being before *B*. Thus one instant *A* is *absolutely before* or *chronologically precedes* another, *B,* if a physical influence can be propagated from *A* to *B*. In such a case, *B* is *absolutely after A*:

> Thus if I can send out any influence or material particle from a particle P at the instant A so as to reach a distant particle Q at the instant B, then this is sufficient to show that B is after and therefore distinct from A. (Robb 1921, 11)

We may call this the Principle of Chronological Precedence, or CP. As can be seen, it presupposes the Principle of Retarded Action discussed in Chap. 4, according to which every physical process takes a finite quantity of time to be completed. Note that so long as CP holds for the propagation of any physical influence, it will not matter whether light or anything else actually travels with the limiting velocity.[41]

As Robb showed in 1914, this means that—restricting temporal relations to these absolute relations only—a given event can be related in order of succession to any event in its future or past light cones, but cannot be so related to any event outside these cones (in what came to be called the event's "Elsewhere"). There are therefore pairs of events that are not ordered with respect to (absolute) before and after, such as the events happening at the instants **A** and **B** on Robb's "Fig. 6.1". The event **B**, being too far away from **A** for any influence to travel between them, is neither before nor after **A**.

---

is another time apart from spacetime; and since one event may chronologically precede another contingently, qualifying it as 'necessarily' earlier is inappropriate.

[41] Robb acknowledges that "it may or may not be strictly true of light" that it travels with this limiting velocity, but he takes light to serve for this purpose in his construction.

For example, **B** could be the event on some planet in the Andromeda Galaxy that Paul Davies asked us to imagine, in the Elsewhere of me at the instant **A** when I am considering it. It is true that by walking this way and that I could describe that event as being in the past or in the future according to the time coordinate associated with the frame of reference in which I am at rest. But that event is not present to me in the sense of being a possible part of my experience. It bears no absolute temporal relation to my considering it. And thus, according to Robb, we cannot say that the two events stand in a relation of succession. Neither is earlier than the other. All the events I experience, on the other hand, will be either before or after one another, and therefore distinct. In fact, they will occur in a linear order. They will lie on what Minkowski called my worldline.

There is nothing unique about my worldline, however. On pain of solipsism, what goes for me goes for any other possible observer (this is the counterpart in his theory to Putnam's "No Privileged Observers").[42] Thus if we regard time as constituted by these absolute relations, time as a whole does not have a linear order: not all events can be ordered on a line proceeding from past to future, even though two events that are in each other's elsewhere (i.e. lying outside each other's cones) will be in the past of some event that is suitably far in the future of both of them. In this way, all events can be temporally ordered, even if not every pair of events is such that one is in the past or future of the other. This is Robb's "conical order". In the language of the theory of relations, it is a strict partial order, rather than a serial order.

One could regard all this as little more than a restatement of what Minkowski had already presented as a consequence of special relativistic spacetime. But Robb's insight was to see that by beginning with this relation of chronological precedence CP so defined, together with postulates defining its properties, the whole mathematical structure of Minkowski's spacetime could be derived without assuming anything else. As we saw in Chap. 4, Leibniz had shown how to construct a theory of temporal order on such a basis, a "causal theory of time" (in which he was followed by several others). But Robb's construction also delivered the affine structure of Minkowski spacetime (basically the system of timelike, spacelike and null lines), as well as grounding congruence between intervals and a metric, all on this foundation. Crucially, he achieved this without making any reference to clocks or measuring rods.[43]

Although Robb's own construction was somewhat daunting,[44] other theorists have picked up on the essence of his approach to give equivalent results using more

---

[42] Thus Robb: "A normal individual who is not a solipsist (and a solipsist could hardly be regarded as a normal individual) believes in the existence of more than his own self and his own perceptions, and one is accustomed to regard these perceptions, under normal circumstances, as representing things as real as one's self but in some sense external" (Robb 1936, 7–8).

[43] Objecting to Einstein's postulating of the existence of clocks to define times Robb wrote "It does not appear a satisfactory mode of procedure to found a philosophical theory upon a complicated mechanism like a *clock* without any precise definition of what constitutes equal intervals of time." (Robb 1914, 8).

[44] "I have been told more than once," Robb ruefully admits in the Preface to *The Absolute Relations of Time and Space*, "that my book [his 1914 *Theory of Time and Space*] is difficult reading." (Robb 1921, vi). There are 21 postulates, the last of which (equivalent to Dedekind's postulate of continuity) finally appears on p. 330! John Winnie gives a thorough examination of Robb's construction in his

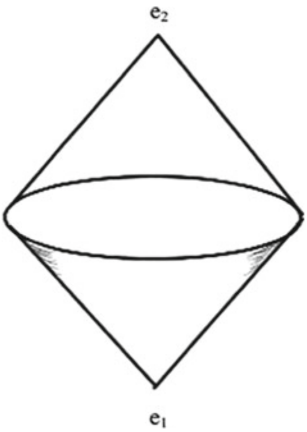


Fig. 6.2 The Alexandrov interval between e_1 and e_2

economical constructions. In a paper of 1967 the Russian mathematician Alexandrov showed how the topology of Minkowski spacetime is uniquely determined "by the propagation of light or, in the language of geometry, by the system of the light cones", noting the equivalence of this derivation to Robb's derivation on the assumption of chronological precedence. The idea is to take as the basic spacetime interval between two point-instants all those point-instants that are absolutely after the earlier instant e_1 and absolutely before the second e_2. This gives the double-cone structure that is now known as the Alexandrov interval (Fig. 6.2).[45]

Other aspects of Robb's theory can be derived with great economy by taking a group-theoretical approach. Thus in his (1964) E. C. Zeeman took such an approach to prove that "Causality Implies the Lorentz Group": that is, that the transformations of spacetime that respect the principle of retarded action are the Lorentz transformations together with rotations in space. Moreover, although Zeeman, like Robb, assumes that the limiting velocity of physical processes is the velocity of light, c, it is possible to construct relativity theory without assuming Einstein's Light Postulate. You need only suppose that there is a finite limiting velocity k, greater than zero (which is infinite in the classical limit). This was first proved algebraically by Vladimir Ignatowski in his (1910), since when essentially the same result has subse-

---

(1977), relating it to those of Alexandrov and Zeeman mentioned below. As he remarks, "Given Robb's ambitious aims, and the conceptual economy of his foundations, the complexity of his constructions was inevitable. … Only after a few hundred theorems do coordinates appear, and with them the Minkowski metric in its standard form." (Winnie 1977, 158).

[45] The *Alexandrov interval* is discussed by Winnie in his (1977, 156–57), with a diagram depicting this interval for two chronologically connectible events. As Winnie explains, Alexandrov intervals have a profound foundational significance, since a topology for Minkowski spacetime may be defined taking these intervals as basis. This is discussed in detail in (Hawking and Ellis 1973, 196ff.), who show that they are also a sufficient basis for defining topologies of spacetimes in general relativity, provided the *strong causality* condition is met. See also (Savitt 2009).

quently been rediscovered several times. In particular, it can be derived very neatly using group theory.[46]

I sketch such a derivation in the Appendix to this chapter. One supposes versions of the principles of inertia and relativity, together with certain natural assumptions about the symmetries appropriate to displacements produced by boosting a frame: namely, linearity, symmetry under space-reflection, an equal times principle, and "causality", here embodied in the requirement that there exist (at least some) time intervals which do not change sign under any boost transformation. This is the constitutive assumption underlying Robb's principle of chronological precedence, CP. It is then proved that there are only two possibilities consistent with these assumptions. Either the "rapidity"—Robb's term for the additive parameter φ characterizing these boosts—reduces to the classical (instantaneous) velocity, $\varphi = v$, and we recover the Galilean transformations; or $\varphi = k \tanh^{-1}(v/k)$, k is determined empirically to equal c, and we obtain the Lorentz transformations.

This derivation of the Lorentz transformations throws light on the status of *tachyons*, putative faster-than-light processes. It is often said that the existence of tachyons is not ruled out by the Lorentz transformations: just as c is a limiting velocity which slower-than-light processes could never achieve, so it is a lower limit for faster-than-light (superluminal) processes. Nevertheless, tachyons are typically dismissed on philosophical grounds: tachyonic processes, it is objected, give rise to causal paradoxes. For a superluminal process going from one spacetime point p to another q extends across a spacelike interval of spacetime. It follows that if it goes from p to q in one inertial frame, so that p causes q and is therefore earlier than q according to the time-coordinate t of that frame ($p <_{t} q$), then one can always find a different inertial frame with time-coordinate t' in which q causes p (q is earlier than p) in that frame, ($q <_{t'} p$). This, it is said, violates "causality", the principle that the cause must always precede the effect. But this way of phrasing the objection proceeds as though causality were some optional add-on to the pre-existing structure of Minkowski spacetime. It is not. The violation of "causality" is more profound than this, and it is worth trying to understand why.

First, we may note the objection of Lévy-Leblond and Provost that transformations from one inertial frame to another are parametrized by rapidity, not velocity. Rapidity, however, ranges between $-\infty$ and $+\infty$, and is not defined for velocities faster than c. Thus "there is no rapidity, i.e., no change of reference frame, associated with superluminal velocities" (Lévy-Leblond and Provost 1979, 1045). So the assumption that there can be a transformation from one superluminal inertial frame to another in Minkowski spacetime is without foundation: such transformations would not be part of the Lorentz group characterizing Minkowski spacetime. But we can say more. In our derivation of the Lorentz transformations "causality" is captured by the requirement that there exist (at least some) pairs of events a and b with a before b,

---

[46] See for example (Lévy-Leblond 1976), (Lévy-Leblond and Provost 1979). Group theoretic approaches are not fashionable among philosophers of physics, since it is known that they are inadequate to characterize the more general spacetimes occurring in the general relativity theory. This explains the preference for intrinsic characterization of spacetimes in terms of manifolds and structures defined on them.

such that no transformation can reverse their order—in Robb's terms, such that *a* is *absolutely before b*. As discussed in Chap. 4, this is not the assumption of causality itself (causation could in principle be instantaneous); rather it is mandated by the principle of retarded action. The point is that a process is something that proceeds from a state or event *a* to a state or event *b*: it cannot do this without taking time, nor can it do it by going from *b* to *a*. In this sense "causality" is a principle of becoming: it determines what counts as a process. And in Minkowski spacetime, only pairs of events in timelike or lightlike separation satisfy this principle. In short, a continuous series of point-events from *p* to *q* in spacelike separation, such as a tachyon is supposed to be, is simply not a process. To regard it as one is to contradict one of the assumptions constitutive of Minkowski spacetime.[47]

Now let us turn to Robb's construal of the present in relativity. For him an instant is a primitive concept; it is an element of time of which I may be directly conscious.[48] The instants of which I am conscious form a linear series, but instants as a whole need not. They can form a strict partial order, where if neither of two instants is before the other, they are not therefore identical. Robb illustrates this possibility using cones, as we saw above; but at this point talk of cones is only illustrative, prior to the derivation of Minkowski spacetime. He then proposes that a necessary and sufficient condition for an instant B to be after an instant A is that it be "be abstractly possible for a person, at the instant A, to produce an effect at the instant B" (Robb 1914, 7). For the producing of an effect Robb is thinking of "an influence or material particle" being sent from A to B. The instants, as noted, are point-instants, and the events happening at them are strictly instantaneous events such as "two particles striking one another" (6). Now if an instant C is a finite distance from A then such an influence must take some finite time (by the Principle of Retarded Action), so that C is necessarily distinct from A. But simultaneous events are those happening at the same instant (6). It therefore follows that "the only events which are really simultaneous are events which occur at the same place" (6).

The punctual present, however, is very problematic. To say that what is compresent with an event (such as a person considering) is merely what shares the very same spacetime point is hardly compatible with our normal intuitions of presentness. It almost collapses Robb's position into a temporal atomism like that of the Sautrãntikas, as mentioned above. But unlike them, Robb did not insist that only observed instants could exist, and in order to ensure that the instants a person could be conscious of would form a continuous series, in his 21st postulate he explicitly posited a Dedekindian continuum of them. But his position is still susceptible to a version of one of Zeno's paradoxes, as we saw in Chap. 2 in association with Russell's views and Grünbaum's ontology of point-events. As was argued there, temporal becoming can no more take place in an instant than can motion. Therefore, if becoming takes

---

[47] Kent Peacock has shown (2018, Chap. 4) that it is possible to derive "Lorentz-like transformations for the superluminal case" (54). But the superluminal "motions" so defined will not only not form a Lorentz group, they will not satisfy CP, so I cannot agree with him that they qualify as possible processes.

[48] "An element of time is called an instant and is to be regarded as a fundamental concept. Of any two elements of time of which I am directly conscious one is after the other." (Robb 1914, 4).

place in the present and all that is present at some point is what is at that spacetime point, then, since there is no becoming in an instant, there is no temporal becoming. As we saw, Russell proposed something like this argument as an objection to becoming. Yet his solution to Zeno's parallel argument against motion is that, although there is indeed no motion in an instant, this does not refute the reality of motion, since this consists in a body's having a different position at a later instant. By parity of reasoning, I argued, although a Zenonian argument shows that a process cannot be composed of point-events, it does not preclude there being a process whereby something becomes different at a later time from what it was at an earlier one.

This last consideration, in fact, points the way to an acceptable construal of becoming, and one that "saves the phenomena" of our experience of the present too. This is achieved by recognizing that becoming occurs over a short (even arbitrarily short) duration; even the event of a person considering or apperceiving another event cannot be strictly instantaneous.

6.5 Compresence and Local Becoming

This leads us naturally into consideration of what might be considered present with respect to such an extended event or process. From a phenomenological perspective, this is hardly a new idea. As has long been recognized, events given in our experience, such as our experience of a snatch of melody, are heard as already laid out in order—in this case, as a succession of notes. This is at odds with the above discussion of relativity, where the events are assumed to be instantaneous, and the instantaneous present assumes the possibility of being consciously aware of them at an instant. Obviously, these are the typical abstractions of a mathematical physicist, as is usually acknowledged by qualifying phrases such as "properly speaking" and "strictly speaking". But these abstractions are also responsible for the gulf between mathematical theory and our palpable experiences of events.

Several authors have proposed that this discrepancy between theory and experience can be bridged by introducing the concept of the *specious present*.[49] We already encountered such a view in our discussion of Bergson and Whitehead in Chap. 3. The idea, as originally developed by William James, is that the present or now as we cognize it in practice is "no knife-edge, but a saddle-back with a certain breadth of its own."[50] By this means, our intuitions of presentness as comprising brief processes

---

[49] See in particular the discussions of Dobbs and Broad (1951), and (Dobbs 1951). Dobbs builds on the speculations of Eddington in (Eddington 1946) about the two-dimensionality of time, as well as Russell's discussion of the paradoxes associated with the specious present in (Russell 1948).

[50] James writes: "In short, the practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time. The unit of composition of our perception of time is a *duration*, with a bow and a stern, as it were—a rearward- and a forward-looking end. It is only as parts of this *duration-block* that the relation of *succession* of one end to the other is perceived." (James 1890, 609–610; quoted from Čapek 1971, 133). Broad (1938, 281ff.) finds others' attempts to define the specious present

and also as encompassing a considerable spatial extent can be preserved. For we do not have to restrict our notion of contemporaneity to what is present to a *point-event* or *instant*, but can apply it to a small extended event of apperception.

Such a construal of contemporaneity has been discussed by Howard Stein in a couple of very influential papers (1968, 1991). The first of these was offered as a rebuttal of the views of Putnam and Rietdijk that we considered above. In a footnote in the first of these Stein provides a definition of contemporaneity that he considers appropriate for Minkowski spacetime in terms of *mutual communication or influence*:

> for processes of more than instantaneous duration, a meaningful and intuitively satisfying notion of 'contemporaneity' can be defined: two such processes may be said to be contemporaneous if part of each is past to part of the other—in other words, if mutual influence ("communication") is possible between them. (Stein 1968, 15 n. 14)

Amplifying on this construal of contemporaneity in terms of interaction in his later article, Stein suggests, as a "plausible anthropological hypothesis", that our "intuitive" notion of the present is grounded in mutual communication, and "first arises 'naturally' in the course of human development and socialization" (Stein 1991, 159). He writes:

> Let us consider a "specious present" π of some percipient being; and let us call an event e "contemporaneous" with π if signals—interaction—influence—can occur *mutually* between e and π. In the Newtonian case, the spatial extent of the set of events contemporaneous with a given specious present is infinite; and it is rather natural to see in this fact the precise correlate, in the physical theory, of the "intuitive" notion of a "present" throughout all of space. (Stein 1991, 159)

Here Stein is linking the notion of mutual communication with what kinds of interaction are countenanced by the physics in question. In the case of Newton's theory, gravitational interaction is assumed to be instantaneous. Kant had already exploited this connection between simultaneity and instantaneous action at a distance in his Third Analogy of Experience. The infinite spatial extent of the hyperplane of simultaneity is grounded in physical interaction. In order for events to be experienced as simultaneous, they must be conceived as being in mutual interaction, thus accounting for the grounding of the instantaneous interactions of Newton's theory.

Perception, however, does not occur by gravitation. So if we are trying to account for how human beings have a conception of the present as including remote events, gravitational action is not to the point. Our intuitions about the present are garnered from what we sense as compresent, and this is dominated by our sense of sight, as noted above in connection with Rømer, Einstein and Robb. Now given the finiteness of the speed of light, and also its value, Stein's suggestion permits us to redeem the notion of the present as having a spatial extent that is, in fact, extremely large. As he explains, it follows that

> the "graining" of time with respect to which a percipient organism can experience conscious interaction with its environment must be such that the "moments" of time (the specious

---

unintelligible, and gives his own. See Davies (1995, 265–278) and Dainton (2010, 103–120) for more contemporary discussions.

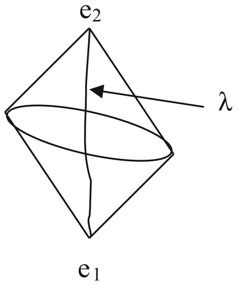


Fig. 6.3 The Alexandrov present, ALEX (e_1, e_2, λ)

> presents) are long enough to allow such signals—and, therefore, light signals—to travel very many times the maximum spatial dimensions of the organism together with its (relevant) environment. (Stein 1991, 161–2)

This explains at one stroke not only why the specious present is of roughly the length it is, but also why humans would naively suppose that the spatial extent of the present—i.e. what we regard as compresent to us during such a specious present—is indefinitely large.

Implicit in this account of Stein's is a notion of the present that can be extended to the relativistic case. For if *a* is the extended event of our becoming aware of some other extended event *e* during some short interval π of proper time, then both *a* and *e* will be short processes, and accordingly both will be represented as segments of world lines in a Minkowski diagram. On this basis, a definition of the extended present appropriate for Minkowski spacetime was independently proposed by myself and Steven Savitt as a natural generalization of Stein's argument. Mine ran:

> the *present* of a process during an interval of its proper time π (for instance, that during which conscious experience is laid down) is that region of spacetime comprised within the absolute future of the beginning of this (usually short) interval and the absolute past of the end of that interval. (Arthur 2006, 148)

In terms of Minkowski spacetime, the absolute future is the forward lightcone, and the absolute past the backward lightcone. The result is that the present of an object during an interval of its proper time is a region of spacetime finite in extent, the intersection of the two "cones", as in Fig. 6.3. Because the region is that which Alexandrov had previously identified as basic to the topology of Minkowski spacetime, Savitt (2009) called it the Alexandrov present, or ALEX (e_1, e_2) for short, where e_1 and e_2 are two events on timelike curve or worldline λ, with e_1 earlier than e_2: "I propose that ALEX (e_1, e_2) is the present for the interval from e_1 to e_2 along λ." (Savitt 2009, 357).[51]

In his (2015) Savitt prefers to call the Alexandrov present a *causal diamond*, since this is the term used in the physics literature for the Alexandrov interval. As

---

[51] Steve Savitt and I discovered we were working along similar lines in an email exchange in 2003. My draft of the paper I was working on then became (Arthur 2006); Steve's draft paper "Time Travel and Becoming" (my copy is dated November 26, 2004) was eventually incorporated into later published papers, like his (Savitt 2009).

he notes, some physicists define a causal diamond as including the boundary of the Alexandrov interval, that is, as the closed set of points in the future of some point p and in the past of another q, while others prefer to exclude the boundary. The above definitions would include processes on the boundary, since a continuous, future-directed lightlike curve from p to q would represent the paths of a process (such as a light ray) going from p to q at the limiting velocity c, and such a process would be absolutely after p and before q. Since such a process could not interact with the process whose worldline extends from p to q, however, it is preferable not to include the boundary in the definition of the Alexandrov present.

Accordingly, in his (2015) Savitt defines his causal diamonds as excluding the boundary. For a spacetime that is "strongly causal" (we'll return to that in Chap. 7) and temporally oriented, then if $I^+(p)$ is the set of points reachable from p by a continuous, everywhere future-directed timelike curve, and $I^-(q)$ is the set of all points from which a continuous, everywhere future-directed timelike curve can reach q, then "the set $I^+(p) \cap I^-(q)$ is a causal diamond" (Savitt 2015, 19, n.3). Savitt discusses the use of these diamonds by theoretical physicists in a variety of contexts, and notes that some cosmologists use diamonds for which the proper time separation between p and q is not small compared to the curvature of spacetime.

Savitt's conception of the present in relativity, however, is not quite the same as mine, and this may have been the source of some confusion in the literature. One thing I omitted above from his definition of the Alexandrov present in his (2009) is that his two events e_1 and e_2 "are 1 s apart on a timelike curve, λ" (2009, 356). Savitt chooses this interval of 1 s for convenience, as "a middling value" between the upper (0.5 s) and lower (3 s) bounds of the specious present suggested by "an eminent psychologist" (359). That would make the present relative to the specious present of human observers. On my construal, on the other hand, the present is indexed to a small section of a worldline that is "usually" of a short proper duration and that *could be* that of "an observer considering", but need not be.[52] For, as I wrote of this construal of the present,

> although it can accommodate what may be present to an observer's conscious experience—and thus preserve our intuitions about the great extent of the present at any moment of consciousness—it does not depend on it. *Any segment of a worldline will have a region of spacetime that is present to it* according to this definition. (Arthur 2006, 148)

According to this construal it makes sense to talk of today or this year as "the present", and to refer to the "present epoch" even in a cosmological context. In each case the present will comprise all those extended events or processes with which the Earth may be said to be in interaction within the proper duration concerned. This is why I called it the "interactive present" in my (2006). Within such a large present all kinds of events can have become and can be in the process of becoming. Savitt, however, "thinks of the present as a locus of becoming, rather than as the 'cutting edge' of what has become" (2015, 23), so that passage is conceived by him as consisting in a

---

[52] Thus Savitt's claim that "Arthur (2006) and I (Savitt, 2009) proposed that the present in (time-oriented) Minkowski spacetime should be thought of as a small causal diamond" (Savitt 2015, 19) is not quite accurate: for me it is not necessarily small.

succession of presents, each one a localized region of spacetime in which becoming occurs. I'll return to this below.

First, however, we should be clear about what purpose the Alexandrov present is intended to serve. Stein's account of presentness in terms of interaction is intended to explain why the present appears to be short in duration and yet large in spatial extent. This depends on the finiteness and the large value of c (in terms of human centred units such as the metre and the second), but not on any features specific to relativity theory. It explains what Whitehead called "presentational immediacy" (Whitehead 1930, 13), if that is understood as the feeling of the immediacy of certain processes experienced as present to us, without assuming (with him) that all processes we regard as present are present to each other in some objective moment. The idea of such an absolute simultaneity is an inferred concept in classical physics, as we saw in Sect. 6.2, and goes beyond what is actually experienced. The virtue of Stein's account is to show how it is not necessary to presume it in order to account for the origins of our intuition of "presentational immediacy". Savitt and I had followed Stein in arguing that the endowing of the present of human awareness with a short duration allows us to account for intuitions of the present being of large spatial extent, and had proposed that an extension of this into relativistic physics required this extent to be the spatiotemporal region given by the Alexandrov interval or causal diamond.

Here it must be stressed that the Alexandrov present is not intended to reproduce all the features of the present of untutored common sense, the phenomenal present. As Mauro Dorato has observed, not only would such a phenomenal present include massively distant events (such as the supernova observed as happening by Chinese astronomers in 1054 CE, mentioned at the beginning of this chapter), it would not include events to which we were not paying any attention, or point-events which were merely possible and not actual.[53] Nevertheless the interactive present is adequate to explain why we perceive certain events happening at a distance (even thousands of kilometres away) as being present, and explains our intuition of the "reality" or robustness of present events in the sense of their being "at hand".[54]

Third, the Alexandrov present is not proposed in order to explain becoming. It presupposes—as is implicit in Robb's work and made explicit by Stein—that becoming takes place along worldlines, and not by the advance of a world-wide now or plane of simultaneity. As Savitt notes, I myself had argued this in an earlier paper (Arthur 1982),[55] and such a conception is endorsed by Dieks, Dorato, Myrvold and

---

[53] Most of these points were made in my article: "The interactive present is not the same, however as the *passive* or *subjective present*, the set of all those events of which we are consciously aware at the moment of considering them" (Arthur 2006, 151). Dorato acknowledges this in his (2011), so his criticism of our proposal on these grounds seems to involve some misunderstanding. Cf. Savitt's response: "Dorato is surely right when he says that causal diamonds, if proposed as a scientific successor concept to our common sense concept of the present, do 'not correctly pick out the events we intend to pick out when we use 'now' in ordinary language.' It is true, however, that nothing in M does." (Savitt 2015, 21).

[54] See Savitt's (2015) for a detailed responses to this and other objections raised by Dorato in his (2011) to the "Arthur-Savitt" conception of the present.

[55] "Suppose we then conjecture the following (Arthur 1982: 107): 'It is this proper time which is understood to measure *the* rate of becoming for the possible process following this timelike line

others.[56] Thus, on the view I am advocating, events come to be in the present in a quite specific sense: if one temporally extended event *b* lies in the Alexandrov present of another *a*, then *b* comes about during the proper time of the event *a*. Such a notion is neither symmetric (even though a part of *a* will lie in *b*'s present, and vice versa) nor transitive. So there is no question of this construal supporting the notion of the present as an equivalence class of events separating the past from the future.

This is relevant to an objection raised by Craig Callender to Stein's conception of the relation of "having become for". He claims that "the thinnest requirement one might put on becoming" is the condition that "at least one event in the universe shares its present with another event's present" (Callender 2000, S592). As we have just seen, this requirement is not satisfied by the above account, where one extended event may share a great deal of its Alexandrov present with another, but cannot share exactly the same present. But Callender is thinking of point-events, and he offers the above characterization as a gloss on a condition that he insists must be satisfied by any relation R*xy*, "*x* has become for *y*", where *x* and *y* are point-events in Minkowski spacetime. This is what he calls the "non-uniqueness condition", NU, defined as $\exists x \exists y \exists R(Rxy \ \& \ Ryx \ \& \ x \neq y)$. Now, Stein's relation B*xy* is provably antisymmetric: that is, for any point-events *x* and *y*, if *x* has become for *y* and *y* has become for *x*, then *x* and *y* are identical. NU simply asserts the existence of events that contradict this property. But why should we accept NU? And why should we believe that NU encodes the property of sharing a present that Callender believes is essential to R?

It would seem that this derives from Stein, who (hypothetically) asserts the equivalence of "sharing a present" with NU in the antecedent of a conditional he states in his article of 1968: "if 'presentness to each other' of events is taken to mean that *each has for the other already become*, then … in Einstein-Minkowski space-time *an event's present is constituted by itself alone*." (Stein 1968, 15). This follows from the anti-symmetry property of his relation B*xy*. But as I understand his argument, Stein is not arguing for the truth of antecedent of the stated conditional statement, but rather that its consequent—namely Robb's construal of the present—is untenable. And granting the negation of its consequent, the conditional entails that compresence (the presentness of events to each other) should *not* be understood to mean that each has for the other already become. Hence the need to adopt a different conception of compresence in terms of the possibility of mutual interaction between temporally extended events. The extended events contained in ALEX (e_1, e_2) will have become for e_2, but not for e_1. They will be becoming while (e_1, e_2) is in process, but will not wholly have become until e_2.

Now if one insists on linking becoming with being in the present, then one could regard the whole region of ALEX (e_1, e_2) as the locus of becoming for the extended event (e_1, e_2). Becoming would consist in a succession of presents or causal diamonds *D* aligned along a worldline. If this is what Savitt means by his characterization of the

---

(or *worldline*)'. It is this idea of becoming along a timelike line, local becoming, that underlies my negative evaluation of the two anti-passage arguments presented above." (Savitt 2011, 11).

[56] See Dieks (2006), Dorato (2006a), Myrvold (2003, Sect. 2).

present as “a locus of becoming” (2015, 23), then his indexing of D to a proper time of about 1 s acquires an ontological significance lacking in my characterization of the present as simply relative to a segment of a worldline. For it implies that becoming takes place in a process in a small region of spacetime corresponding to 1 s of its proper time, but that events cannot be said to have become until after that proper time has elapsed. Events are in the process of becoming in each D, but nothing has become until after each D.[57] This is somewhat evocative of Bergson and Whitehead, for whom the continuous, extensive, linear time of mathematical physics is an after-effect of real becoming (*la durée réelle*), which happens in “indivisible, temporal pulsations”.[58] But Bergson’s *la durée* is “immediately perceived duration”, and we have said enough about the mistake of attempting to base physical time on subjective experience, which is no part of my or Savitt’s remit.

Turning to physical time, as opposed to the time of psychology, the interval of 1 s appears arbitrary, as Dorato has objected (2011, 302). But the examples Savitt gives of the use of causal diamonds in contemporary physics suggest that a finite time-indexed present might have application at much smaller time scales, where quantum considerations would come into play. There a discrete temporal extent is motivated, not by considerations of its relationship to human perceptions, but by considerations of the indeterminateness of becoming below a certain threshold. But before we enter into quantum considerations, we have first the small task of digesting Einstein’s second revolution in our concepts of space and time, wrought by his theory of General Relativity!

As I shall try to show, however, there are problems in the interpretation of both general relativity and quantum theory that arise from an inadequate appreciation of the lessons of special relativity. As Savitt has said, the most important and surprising of such lessons is that temporal passage “is a *local* phenomenon, tied to a world line”, and not tied “to an advancing *global* now”. It is the latter idea, “buried deep in our worldview, … that we must transcend” (Savitt 2011, 14).

6.6 Summary

- This chapter treated the topic of the status of the present in the face of the relativity of simultaneity of SR. Beginning with a discussion of the status of the present in classical theory, I argued that the events appearing in phenomenal experience are all events that occurred in the past. This already creates difficulties for the kind of primacy accorded to the events of present experience, as in Whitehead’s

---

[57] Savitt provides a definition of becoming relative to a causal diamond D in his paper (2015, 23).

[58] Here I am quoting Milič Čapek’s synthesis of the views of Bergson and James: “Mathematical continuity is thus nothing but *discontinuity endlessly repeated*, which ignores the natural articulation of duration into indivisible, temporal pulsations” (Čapek 1971, 139). This is similar to Whitehead’s atomic becoming, discussed in Chap. 2. Bergson himself characterizes *la durée* as “uninterrupted transition, multiplicity without divisibility and succession without separation” (Bergson 1965, 44–45).

idea of “presentational immediacy”, and necessitates a distinction between the phenomenal present and the inferred present. An easily visualizable illustration of the distinction is given using the observations of time that would be made by the twins of SR’s Twin Paradox according to classical theory.

- One of the main responses to the difficulty of the relativity of simultaneity in SR has been to cast the present as *observer-dependent*. Insofar as this involves conceiving the universe as a static block and making becoming wholly subjective, this has already been treated in Chap. 3. But a second contributing factor to this view has been the idea that observers have their own reference frames, and that time lapses differently for them because they see different events as present. In order to untangle the misconceptions here, I continued the illustration of the twin paradox to show what the twins would perceive, what they would infer, and how this is all mutually consistent, without assuming any observer-dependence of becoming. This reinforces the conclusion of Chap. 5 that time lapse is perfectly objective in SR, and measured by the invariant proper time, and not by the frame-relative time coordinate function.
- Consequently the temporal ordering of point-events in Minkowski spacetime is not a total ordering, but a partial ordering, which A. A. Robb in 1911 termed a “conical order”: not every two point-events occur in an order of succession. Rather all and only those events that are connectible by a physical influence are earlier or later—Robb’s postulate of Chronological Precedence (CP). From this Robb inferred that there was no temporal relationship between events in space-like separation, and proposed that an event’s present does not extend beyond the point of spacetime where it occurs. It was also explained how, from this postulate of CP together with the principle of relativity and certain natural constraints on physical process, Robb was able to construct Minkowski spacetime; a result that was fortified by the contributions of later authors. There is a perfectly well defined spatiotemporal interval between two events that are space-like separated, but no time interval. It was argued that this perspective on SR precludes regarding tachyons as processes.
- Robb’s punctual present is hard to square with the robustness of our sense of the present as having some extension. I suggest a way of defining a notion of the present that is compatible with special relativity, and also with the temporal extendedness of events, such as that of experiencing an event. Such a present is relative to segment of a worldline, and is comprised by a four dimensional region of Minkowski spacetime contained within the two hypercones centred on that segment. I argued that such a notion could account for the compresence of certain enduring things, although it would not be identical with either the phenomenal present or the inferred present of classical physics.

Appendix: Informal Sketch of a Group-Theoretic Derivation of the Lorentz Transformations[59]

The basic idea is this. Inertial frames are characterized not by rods and clocks, but by Lange's construction (Lange 1886), where bodies moving relatively to one another in straight lines without circular motion produce equal displacements in equal times. Such motions form an equivalence class, and this class defines an inertial frame. The Principle of Inertia is then constituted by the assertion that there exists at least one such inertial frame. The Principle of Relativity, in its general form, then asserts the equivalence of these inertial frames for expressing the laws of physics; more precisely, that there exists an infinite class of equivalent inertial frames in relative motion one to another, such that the laws of mechanics are invariant under transformations from one inertial frame to any other. The idea is to derive the various possible forms of this group of transformations by imposing certain plausible constraints on them. A succinct way of achieving this that brings out some essential features of Robb's construction may be summarized as follows.

The same vector displacement that is produced by a point moving with a uniform velocity in a given inertial frame (which we may term, conventionally, the "stationary frame") may equivalently be produced by a point at rest in a frame moving with the same velocity relative to the original inertial frame, that is, by an active boost in velocity of that frame. We may call this a "boost displacement". The Principle of Relativity can now be understood in group theoretical terms as asserting that these transformations by a boost in uniform velocity v, constitute a group of transformations, more precisely, a differentiable and connected one-parameter group. Now, according to a standard mathematical theorem of group theory, for any such group there exists an *additive parameter,* such as the *angle* for a group of *rotations* in Euclidean space. This parameter must be a function of the velocity, and also have the dimensions of velocity. Classically, it was assumed that it simply *is* the velocity. But given the fact that lengths and times in moving frames are not invariant (due to length contraction and time dilation), we cannot suppose that this holds universally. So as not to prejudge things, following Robb we will call this additive parameter producing displacements by velocity boosts the "rapidity", denoted φ.[60] The additive nature of the parameter means that the rapidity addition law is just like the velocity addition law in classical mechanics, $\underline{\varphi}_{12} = \underline{\varphi}_1 + \underline{\varphi}_2$. It is the analogue of the Parallelogram Law, $\underline{\mathbf{v}}_{12} = \underline{\mathbf{v}}_1 + \underline{\mathbf{v}}_2$, expressed by Newton as follows:

---

[59] This sketch is based on (Arthur 2007), which in turn was largely inspired by Provost (1980). See these papers for further details.

[60] "We propose now to define what we shall call the rapidity of a particle with respect to a system of permanent configuration … we shall see later that, for large values, it is the rapidity and not the velocity which follows the additive law." (Robb 1911, 8–9).

> A body acted on by [two] forces acting jointly describes the parallelogram in the same time in which it would describe the sides if the forces were acting separately. (Newton 1999, 417)[61]

Here Newton is assuming that each "force" will produce a corresponding inertial motion proportional to its velocity, and that each inertial motion will produce a displacement in a given time proportional to that time.

Now by the Principle of Relativity any displacement $\underline{\mathbf{x}}$ in any one inertial frame can be realized by an active boost of a point-particle at rest in a second inertial frame in motion relative to the first with a rapidity $\underline{\varphi}$. What plausible constraints should there be on these boosts beyond the rapidity addition law? One of these is linearity, which can be shown to follow from the *homogeneity* of spacetime. Second, the boost displacement should obey *space-reflection*, so that that the relative rapidity $\underline{\varphi}$ of two inertial frames changes sign under change of sign of all the spatial axes: if $\underline{\mathbf{x}} \to -\underline{\mathbf{x}}$, $\underline{\varphi}(-\underline{\mathbf{x}}) \to -\underline{\varphi}(\underline{\mathbf{x}})$. Third, they should obey the same *equal times principle* as that holding for displacements in the stationary frame according in Newton's Parallelogram Law: if three boost displacements $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \underline{\mathbf{x}}_3$ satisfy $\underline{\mathbf{x}}_1 + \underline{\mathbf{x}}_2 = \underline{\mathbf{x}}_3$, then the displacements on both sides of the equation have been realized in equal times. Fourth, we assume *causality*, which here consists in the requirement that there are (at least some) time intervals which do not change sign under any boost transformation. This last constraint rules out a group of transformations which are such that one may always find a boost that changes the sign of the time interval. In Galilean spacetime, all time intervals satisfy this constraint; in Minkowski spacetime, it is satisfied by pairs of events separated by a timelike or lightlike interval. So in those cases the constraint is therefore equivalent to Robb's assumption that there are asymmetric relations of chronological precedence that are absolute (independent of frame).

Expressing these constraints in group theoretical terms, one can prove that there are only two possibilities consistent with all of them (which we can express in one dimension without loss of generality). The first just returns $\varphi = v$, rapidity is identical with velocity, as was assumed classically. From this we obtain the Galilean transformations consistent with the classical velocity addition law, $\underline{\mathbf{v}}_{12} = \underline{\mathbf{v}}_1 + \underline{\mathbf{v}}_2$. The second and more interesting case delivers $\varphi = k \tanh^{-1}(v/k)$, or equivalently, $v = k \tanh(\varphi/k)$, with rapidity adding according to $\varphi_{12} = \varphi_1 + \varphi_2$. Here k is an undetermined constant with the dimensions of velocity, whose value is determined experimentally to be equal to c, the speed of light in a vacuum. Notably, this does not need to depend on anything to do with light. It would suffice, for instance, to determine that at a relative velocity of 3/5 of the value of c there is a time dilation of some process by a factor of 5/4. Taking units where $c = 1$, as is customary, rapidity is given by $v = \tanh \varphi$, in agreement with Robb's definition.[62] As he explained there, this gives the Lorentz transformations in the following succinct form:

---

[61] A more literal translation would be: "A body [carried] by conjoined forces describes the diagonal of a parallelogram in the same time as [it would] the sides by the separate forces." Newton assumes that the forces and velocities are directed, so that his law is a law of vector addition, *avant la letter*. For simplicity's sake, we may consider it in one dimension only without loss of generality.

[62] "If v be the absolute velocity of the particle with respect to the system, then the inverse hyperbolic tangent of v will be spoken of as the rapidity. Thus if ω be the rapidity, $v = \tanh \omega$. As ω increases

$$x' = x \cosh \varphi - t \sinh \varphi \quad \text{—equivalent to } x' = \gamma(x - vt), \gamma = 1/\sqrt{\left(1 - v^2\right)}$$
$$t' = t \cosh \varphi - x \sinh \varphi \quad \text{—equivalent to } t' = \gamma(t - xv), \gamma = 1/\sqrt{\left(1 - v^2\right)}$$

What we have, then, is a bifurcation of the classical velocity concept into two distinct concepts: the *velocity* in a given inertial frame of reference, and the *rapidity* characterizing active transformations consisting in boosts in instantaneous velocity (or, equivalently, passive transformations from one inertial frame to another in relative motion to it).[63] We may say that the classical concept of velocity is degenerate, and that this degeneracy is only revealed as velocities become an appreciable fraction of c. This is analogous to the degeneracy of classical time noted in the previous chapter, where the classical time concept that did double duty in keeping track of synchrony and tracking becoming, at velocities close to c bifurcates into co-ordinate time (keeping track of synchrony) and proper time (tracking becoming).

References

Alexandrov, A.D. 1967. A Contribution to Chronogeometry. *Canadian Journal of Mathematics* 19: 1119–1128.

Arthur, Richard T.W. 1982. Exacting a Philosophy of Becoming from Modern Physics. *Pacific Philosophical Quarterly* 63 (2): 101–110.

Arthur, Richard T. W. 2006. Minkowski Spacetime and the Dimensions of the Present. In Dieks, ed. 2006, 129–155.

Arthur, Richard T. W. 2007. Time, Inertia and the Relativity Principle, philsci-archive.pitt.edu/3660/1/Tirp.rtwa5.pdf.

Bain, Jonathan. 1998. Whitehead's Theory of Gravity. *Studies in the History and Philosophy of Modern Physics* 29: 547–574.

Bell, J.S. 1987. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Bergson, Henri. 1944. *Creative Evolution*. Trans. *L'Évolution créatrice*, 1907. New York: Henry Holt and Company, originally published 1911.

Bergson, Henri. 1965. *Duration and Simultaneity*. Trans. *Durée et simultaneité* (1922) by Leon Jacobson, with an introduction by Herbert Dingle. Indianapolis: Bobbs Merrill.

Borges, Jorge Luis. 1964. *Labyrinths.* In Donald A. Yates and James E. Irby. New Directions Publishing.

Born, Max. 1922. *Einstein's Theory of Relativity*. New York: E. P. Dutton & Co.

Bridgman, Percy W. 1927. *The Logic of Modern Physics*. New York: MacMillan.

Broad, Charlie Dunbar. 1938. *Examination of McTaggart's Philosophy*, vol. 2. London: Cambridge University Press.

---

from 0 to ∞, v increases from 0 to 1. For small values of ω we have, practically, velocity is equal to rapidity, but we shall see later that, for large values, it is the rapidity and not the velocity which follows the additive law." (Robb 1911, 8–9).

[63] As Lévy-Leblond and Provost (1979, 1045) remark, it shows the need "to replace the Galilean velocity by two separate concepts: 'velocity' v, as expressing the time rate of change of position, and 'rapidity' φ, as the natural additive group parameter."

Callender, Craig. 2000. Shedding Light on Time. *Philosophy of Science* 67 (Supplement 3): 587–599.

Čapek, Milič. 1966. Time in Relativity Theory: Arguments for a Philosophy of Becoming. In *Voices of Time*, ed. J.T. Fraser, 434–454. New York: Brazilier.

Čapek, Milič. 1971. *Bergson and Modern Physics*. Dordrecht: D. Reidel.

Carmichael, R.D. (ed.). 1927. *A Debate on the Theory of Relativity*. Chicago/London: Open Court.

Dainton, Barry. 2010. *Time and Space*, 2nd ed. Montreal: McGill-Queens University Press.

Davies, Paul C.W. 1995. *About Time*. New York: Touchstone.

Dieks, Dennis. 2006. Becoming, Relativity and Locality. In *The Ontology of Spacetime*, ed. Dennis Dieks, 157–176. Amsterdam: Elsevier.

Dingle, Herbert. 1939. The Relativity of Time. *Nature*, November, 888–890.

Dingle, Herbert. 1965. Introduction, pp. xv–xlii in (Bergson 1965).

Dingle, Herbert. 1972. *Science at the Crossroads*. London: Martin, Brian and O'Keefe.

DiSalle, Robert. 2006. *Understanding Spacetime: The Philosophical Development of Physics from Newton to Einstein*. Cambridge: Cambridge University Press.

Dobbs, H. A. C. 1951. The Relation Between the Time of Psychology and the Time of Physics, Part 2, 2(7): 177–192.

Dobbs, H. A. C. and C. D. Broad. 1951. The Relation Between the Time of Psychology and the Time of Physics, Part I, 2(6): 122–141.

Doppler, Christian. 1842. *Über das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels* (On the coloured light of the binary stars and some other stars of the heavens). A facsimile edition is contained in (Eden 1992).

Dorato, Mauro. 2006a. The Irrelevance of the Presentist/Eternalist Debate for the Ontology of Minkowski Spacetime. In Dieks ed. 93–109.

Dorato, Mauro. 2006b. Absolute becoming, relational becoming and the arrow of time: Some non-conventional remarks on the relationship between physics and metaphysics. *Studies in History and Philosophy of Modern Physics* 37: 559–576.

Dorato, Mauro. 2011. The Alexandroff present and Minkowski spacetime: Why it cannot do what it has been asked to do. In: *Explanation, Prediction, and Confirmation: The Philosophy of Science in a European Perspective,* vol. 2, ed. Dennis Dieks, S. Gonzalez, T. Hartmann, Uebel and M. Weber. Dordrecht, Heidelberg, London, New York: Springer. http://dx.doi.org/10.1007/978-94-007-1180-8.

Eden, Alec. 1992. *The Search for Christian Doppler*. Wien: Springer.

Eddington, Arthur S. 1946. *Fundamental Theory*. Cambridge: Cambridge University Press.

Einstein [1905] 1923. Zur Elektrodynamik bewegter Körper. *Annalen der Physik* 17(1): 891–921 (translated into English in Lorentz et al. 1923, 37–65).

Einstein, Albert. 1954. *Ideas and Opinions*. New York: Crown Publishers.

Fokker, A. D. 1953/1955. Albert Einstein, Inventor of Chronogeometry. *Synthese* 9 (6A/6B): 442–444.

Gassendi, Pierre. 1649. *Animadversiones in Decimum Librum Diogenis Laërtii* (Critical Remarks on the Tenth Book of Diogenes Laërtius). Lyons: Guillaume Barbier.

Greene, Brian. 2004. *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*. New York: Alfred A. Knopf.

Hawking, Stephen W. 1988. *A Brief History of Time*. London: Bantam Press.

Hawking, S.W., and G.F.R. Ellis. 1973. *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Howard, Don. 2017. Einstein's Philosophy of Science, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/einstein-philscience/.

Hurley, Patrick. 1986. Time in the Earlier and Later Whitehead. In *Physics and the Ultimate Significance of Time*, ed. David Ray Griffin, 87–109. Albany: State University of New York Press.

Ignatowski, W.V. 1910. Einige allgemeine Bemerkungen zum Relativitätsprinzip. *Phys. Zeits.* 11: 972.

James, William. 1890. *The Principles of Psychology*. New York: Henry Holt and Company.

Lange, L. 1886. *Die geschichtliche Entwicklung des Bewegungsbegriff und ihr ausschichtliches Endergebnis*. Leipzig.

Lévy-Leblond, Jean-Marc. 1976. One more derivation of the Lorentz transformations. *American Journal of Physics* 44 (3): 271–276.

Lévy-Leblond, Jean-Marc, and Jean-Pierre Provost. 1979. Additivity, Rapidity, Relativity. *American Journal of Physics* 47 (12): 1045–1049.

Lovejoy, Arthur O. 1931. The Paradox of the Time-Retarding Journey. *The Philosophical Review* 40 (48–68): 152–167.

Marder, L. 1971. *Time and the Space-Traveller*. Philadelphia: University of Pennsylvania Press.

Mermin, N.David. 2005. *It's About Time: Understanding Einstein's Relativity*. Princeton and Oxford: Princeton University Press.

Minkowski, Hermann. 1907. The Fundamental Equations for Electromagnetic Processes in Moving Bodies. Trans. Vesselin Petkov, pp. 51–110 in (Minkowski 2012).

Minkowski, Hermann. 1908. Space and Time. Trans. Vesselin Petkov, pp. 111–125 in (Minkowski 2012).

Minkowski, Hermann. 2012. *Space and Time: Minkowski's Papers on Relativity*. In: ed. Vesselin Petkov. Montreal: Minkowski Institute Press.

Myrvold, Wayne. 2003. Relativistic quantum becoming. *British Journal for Philosophy of Science* 54: 475–500.

Newton, Isaac. 1999. *The* Principia*: Mathematical Principles of Mathematical Philosophy*. Ed. and Trans. I. Bernard Cohen and Anne Whitman. Berkeley: University of California Press.

Peacock, Kent A. 2018. *Quantum Heresies*. Milton Keynes, UK: College Publications.

Provost, Jean-Pierre. 1980. A Truly Relativistic Approach of the Concept of Time. In *Group Theoretical Methods in Physics*, vol. 135, ed. K.B. Wolf, 456–458., Lecture Notes in Physics Berlin/Heidelberg: Springer.

Putnam, Hilary. 1967. Time and Physical Geometry. *Journal of Philosophy*, 64(8), April 27, 240–247.

Robb, Alfred A. 1911. *Optical Geometry of Motion: A New View of the Theory of Relativity*. Cambridge: W. Heffer and Sons; London: Simpkin Marshall & Co.

Robb, Alfred A. 1914. *A Theory of Time and Space*. Cambridge: The University Press.

Robb, Alfred A. 1921. *The Absolute Relations of Time and Space*. Cambridge: The University Press.

Robb, Alfred A. 1936. *Geometry of Time And Space*. Cambridge: The University Press.

Rømer, M. and I. Bernard Cohen. 1940. Roemer and the First Determination of the Velocity of Light (1676). *Isis,* **31**(2) (Apr., 1940): 327–379.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Rugh, Svend E., and Henrik Zinkernagel. 2009. On the Physical Basis of Cosmic Time. *Studies in History and Philosophy of Modern Physics* 40: 1–19.

Russell, Bertrand. 1948. *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.

Ryckman, Thomas A. 2014. Early Philosophical Interpretations of General Relativity. *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2014/entries/genrel-early/.

Savitt, Steven F. 2009. The Transient *nows*. In *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony*, ed. W. Myrvold and J. Christian, 339–352. Berlin: Springer.

Savitt, Steven F. 2011. Time in the Special Theory of Relativity. In *The Oxford Handbook of Philosophy of Time*, ed. Craig Callender, 1–20. Oxford: Oxford University Press.

Savitt, Steven F. 2015. I ✠ ◆s [I Love Diamonds]. *Studies in History and Philosophy of Modern Physics* 50: 19–24.

Shcherbatskoy, Fyodor. 1932. *Buddhist Logic*, vol. 1, (reprinted New York: Dover, 1962, Osnabrück: Biblio Verlag, 1970).

Sklar, Lawrence. 1974. *Space, Time, and Spacetime*. Berkeley: University of California Press.

Stein, Howard. 1968. On Einstein-Minkowski Space-Time. *Journal of Philosophy* 65 (1): 5–23.

Stein, Howard. 1970. A Note on Time and Relativity Theory. *Journal of Philosophy* 67 (9): 289–294.

Stein, Howard. 1991. On Relativity Theory and the Openness of the Future. *Philosophy of Science* 58 (2): 147–167.

Weinstein, Galina. 2012. Einstein's Clocks and Langevin's Twins. https://arxiv.org/pdf/1205.0922.pdf (accessed March 14, 2017).

Whitehead, Alfred North. 1920. *The Concept of Nature*. Cambridge: Cambridge University Press; reissued Mineola, NY: Dover Publications, Inc., 2004.

Whitehead, Alfred North. 1925. *Science and the Modern World*. New York: Macmillan.

Whitehead, Alfred North. 1930. *Process and Reality*. New York: Macmillan.

Whitrow, G.J. 1988. *Time in History: Views of Time from Prehistory to the Present Day*. Oxford/New York: Oxford University Press.

Winnie, John. 1977. The Causal Theory of Space-Time. In *Foundations of Space-Time Theories*, ed. John Earman, Clark Glymour, and John Stachel, 134–205. Wisconsin: Minnesota University Press.

Zeeman, E.C. 1964. Causality Implies the Lorentz Group. *Journal of Mathematical Physics* 5: 490.

Chapter 7
Time in General Relativity

Relativistic clocks are hodometers of timelike worldlines.
—Roberto Torretti, *Relativity and Geometry.**

7.1 Introduction

In his *Letters on the English Nation* of 1733, Voltaire was intent on twitting his compatriots by singing the praises of the experimental science he saw being practised in England. One of the examples he relates concerns Francis Bacon: "We must, says he, make an Experiment to see whether the same Clock will go faster on the Top of a Mountain or at the Bottom of a Mine" (Voltaire 1733, 91). I won't hazard a guess as to what prompted Bacon to ask this question. But to a modern reader familiar with Einstein's General Theory of Relativity the question is striking for its prescience. For it is one of the predictions of that theory that the clock at the top of the mountain would indeed run faster. Although the clocks available to Bacon would not have been remotely able to detect the difference in rates, modern experiments have confirmed Einstein's prediction.[1]

This intimate connection of time with gravity was expected by no one (except perhaps Bacon). Known as gravitational time dilation, the phenomenon is manifested in gravitational red shift, the lowering of the frequency (reddening) of light reaching us from distant stars (since we are, as it were, at the top of a mountain with respect

---

*Torretti (1996, 96). In this quotation I have silently corrected Torretti's spelling of 'hodometer' as 'hodometre'. A hodometer is a device consisting of a wheel attached to a dial (an analogue odometer). Another name for it is a *waywiser*; the term Harvey Brown uses in his (2005), and an illustration of the device is shown on the cover of his book.

[1] This is a very small effect. If the mountain is Ben Nevis at about 1500 m, and the deepest mine is about the same depth, a rough calculation gives the clock at the top of the mountain running faster by a factor of about $1 + 3 \times 10^{-12}$! That is, it would gain by a few microseconds per year over the one in the mine.

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_7

to the star's gravitational field). As we'll see, Einstein himself had deduced this connection between time and gravity as a consequence of his Equivalence Principle in 1907—a principle he had inferred very elegantly from a simple thought experiment. Noting that a body in free fall would experience no forces, just like a body moving inertially, and that the effects on a body undergoing an acceleration of 1 g would be indistinguishable from those on a body in a linear gravitational field, Einstein proposed that the effects of gravity and acceleration—weight and inertia—are not just empirically indistinguishable, but identical. He then showed how gravitational time dilation follows from this as a necessary consequence. But it took him a further eight years of excruciatingly hard work to formulate the field equations that constitute the core of the theory of General Relativity (hereafter, GR), equations from which the quantity of time dilation could be more precisely computed. In the same paper of 1907 Einstein had made a second inference from his Equivalence Principle that was equally if not more sensational: spacetime itself must be curved. Working out how to express this mathematically—given the other philosophical principles and physical constraints he maintained—is what took him so many years to accomplish.

Because of its role in guiding Einstein in his formulation of General Relativity, the Equivalence Principle can be used to explain some of the most innovative features of GR without our becoming embroiled in all the more mathematical intricacies of the theory itself. Some of the most significant of these for the status of time are discussed in Sect. 7.2. Gravitational time dilation is obviously one such feature. But the idea that spacetime itself can be warped has further profound implications for time. One such implication emerges in connection with the phenomenon of black holes, whose existence was a predicted consequence of GR, even though they were initially resisted as unphysical. For an object falling into such a hole, (proper) time will pass normally, but from the perspective of an observer farther away from the hole all the processes going on in it will have slowed to a stop. A second bizarre consequence of GR that was realized while Einstein was still alive was that his field equations allowed solutions where spacetime could be curved back onto itself. In such a spacetime a worldline could curl into a closed loop, apparently making time travel into one's own past in principle possible. As noted in Chap. 5 above, Gödel argued that the very possibility of such closed time-like curves (CTCs) refutes the reality of becoming. I will postpone discussion of this matter to Sect. 7.5, where we will consider issues concerning cosmology and cosmic time.

First, however, we need to be clearer on precisely what General Relativity is—its constitutive assumptions, as well as the interpretive issues it raises. This is the topic of Sect. 7.3. We begin with a summary of the philosophical and physical assumptions guiding Einstein in his tortuous path to the final form of the theory. The chief guiding principle was Einstein's philosophical commitment to the relativity of motion, and the aim to extend it to all reference systems, not just inertial ones. This was his motivation for the Equivalence Principle. As we have seen, he followed Huygens and Mach in insisting on the relative nature of all motion, and interpreted this as a relativity to reference frames conceived as co-ordinate systems defined by material rods and clocks. It was this interpretation that led him to believe that if the equations of his theory could be put into a form that does not depend on the coordinate system adopted—if they could be made generally covariant—this requirement would guarantee such relativity. A third guiding principle follows from the Equivalence Principle as Einstein conceived it, namely, as meaning that every acceleration could be understood as a produced by an equivalent gravitational field. In order for this to be the case universally, any such gravitational field would have to have massive bodies as its source: the metric of the field would have to be, as Einstein wrote in 1918, "exhaustively determined by the masses of bodies". This "is the idea that Mach expressed," he wrote in 1922, "that inertia depends on the mutual action of bodies" (Einstein [1922] 1956, 107); in 1918 he dubbed this Mach's Principle.

This program did not go smoothly. None of Einstein's principles delivered the universal relativity of all motion that he had intended to capture, covariance did not have the physical interpretation he had supposed, and the status of Mach's Principle in the final theory is moot. Nevertheless, Einstein was successful, unprecedentedly successful, in producing a new theory of gravity that merged together inertia and gravity into one unified field, and at the same time revolutionized (again) our understanding of space and time.

But the difficulties with Mach's Principle are best understood by reference to the example that Einstein used to introduce it, that of rotational motion. This is the subject of Sect. 7.4. Here we examine the implications of the failure of rotational motion to conform to the universal relativity of motion, but also the further ramifications for understanding time. For in GR there is no global reference frame associated with a rotating observer. For us on the spinning Earth, for example, there is no global 'now' that we could consistently apply to all distant events, despite the local passage of time. This reinforces the lesson of Chap. 5 above: becoming is local, and not associated with the time coordinate of an observer's proper reference frame.

In Sect. 7.5 we turn to issues about time and becoming in relativistic cosmology. A further momentous implication of Einstein's field equations, though one he resisted for a while, is that spacetime is dynamic: it must either expand or contract. The discovery by the observational astronomer Edwin Hubble that all distant galaxies appear to be moving apart at a velocity proportional to their distance confirmed that the observable universe is in fact expanding. The Belgian priest Georges Lemaître was the first to point out that a class of solutions of the Einstein Field Equations predicted just such an expansion (although the same mathematical result had been obtained earlier by the Russian theorist Alexander Friedmann), and this conformity of Einstein's theory with observation was confirmed and publicized by Arthur Eddington. Assuming the universe has always expanded at its current rate allows us to extrapolate the expansion of spacetime backwards to its most concentrated form, the time of the hypothesized Big Bang, and it is a modified version of this argument that allows us to say that that event took place some 13.8 billion years ago. But like Einstein's static model, the dynamic models involve a time function, cosmic time, during which the spatial extent of the universe either stays the same or increases. The astrophysicist James Jeans suggested that this amounts to the reinstatement of a kind of absolute time by the back door, as it were, with the whole universe evolving in step with cosmic time. As we shall see, however, this cosmic time cannot be used to ground the idea of a universal becoming distinct from the local becoming of individual processes. The principle of local becoming, on the other hand, is embodied in the geodesic principle of GR, which guarantees the same connection of time with inertia as was ensconced in Newton's physics and preserved in SR.

7.2 The Equivalence Principle and Curved Spacetime

We will begin by considering the Equivalence Principle. This is a heuristic principle, rather than a constitutive axiom of GR. It guided Einstein as a kind of fixed point of interpretation, and has continued to serve the same purpose for investigating the physical consequences of the theory. This is indeed fortunate, since the equations of the theory are non-linear, and impossible to solve in the absence of some strongly idealized simplifying assumptions. Yet, as already noted, some of the main revolutionary features of GR, such as gravitational time dilation and the warping of spacetime, are already consequences of the Equivalence Principle, so they can be discussed prior to considering the mathematical theory.

In the classical Newtonian theory of gravitation, the accelerative force on any one mass m due to the gravitational attraction of another mass M—such as the force that keeps the Earth in orbit around the Sun—is proportional to both m and M. The Inverse Square Law is

$$F_G = GMm/r^2 \tag{7.1}$$

where G is Newton's gravitational constant ($6.67{,}408 \times 10^{-11}$ m^3 kg^{-1} s^{-2}). Here m is called the *passive* gravitational mass, and M the *active* gravitational mass. The proportionality to m seems natural enough, since we are familiar with the fact that weight, the action of gravity, is proportional to mass. Thus, in keeping with Newton's Second Law, $F = ma$, we unhesitatingly represent weight as $W = mg$, where W is the weight, m the mass, and g the acceleration due to gravity. But the m in $F = ma$ is the *inertial mass*: it represents the body's resistance to change of its state of motion. Why should it be identical to the mass occurring in the inverse square law, the (passive) *gravitational mass*? On what grounds can we say that the inertial mass (which we may write m_i), is equal to the gravitational mass (m_g), that $m_i = m_g$?

To clarify this point, we may compare with another inverse square law that was modelled after Newton's, Coulomb's law for the electrical attraction between two opposite charges, say q_1 and $-q_2$:

$$F_E = k_e q_1 q_2/r^2 \tag{7.2}$$

where k_e is Coulomb's constant (8.99×10^9 N m^2 C^{-2}). By analogy with this equation we could regard m and M in Eq. (7.1) as "gravitational charges". Why should they be the same as the inertial masses? The acceleration due to the electrical force will be $F_{E/m}$, but that due to gravity will be $F_{G/m} = GM/r^2$, an expression in which m

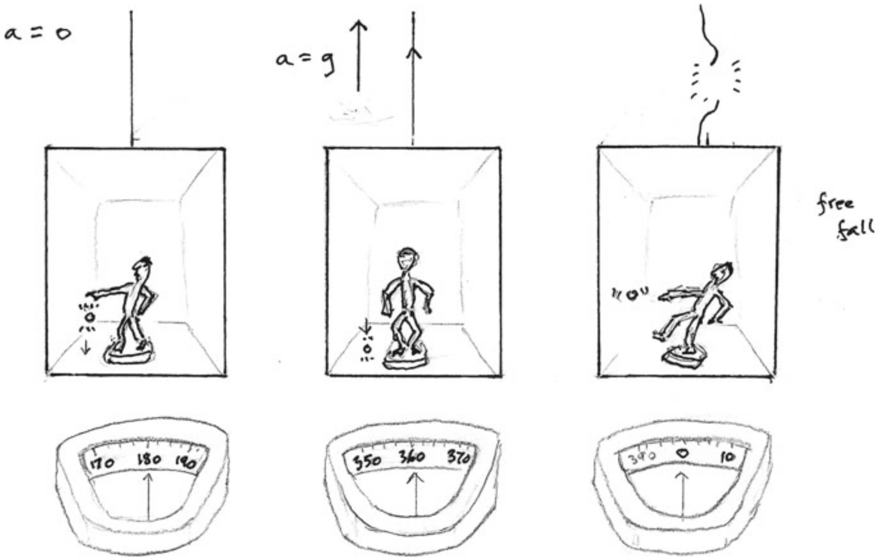


Fig. 7.1 Einstein's happiest thought

does not occur. That is, the acceleration undergone by the Earth due to the Sun's gravitational pull does not depend on the mass of the Earth at all: any body at the same distance from the Sun will experience the same acceleration. This persuaded Einstein that there is no difference: gravity is, in fact, simply an acceleration. The inertial mass m_i is not just equal to the gravitational mass m_g, they are identically the same thing.

Einstein illustrated this with a thought experiment When you are in an elevator, you feel a slight increase in weight as it starts to ascend. By the same token, if the cable were to snap, you would feel no weight at all. "If a person falls freely," he reasoned, "he will not feel his own weight". Not only that: all the objects around you would remain at rest relatively to you. According to Einstein's recollections, this insight—which he later described as "the happiest [*glücklichste*] thought of my life" (CPAE 7, 31; Isaacson 2007, 145)—occurred to him while he was working in the patent office in Bern in 1907. On this way of looking at it, your feeling of weight is due to the fact that you are being held up against a natural tendency to undergo free fall by the floor of the elevator, or by the ground you walk on, or by whatever else keeps you up (Fig. 7.1).

Now in Newtonian theory this is simply a contingent fact. If you feel twice as heavy as your all-too-powerful elevator carries you upwards, then this is because the force exerted on you by the elevator just happens to equal your weight, because its acceleration a happens to equal g and because your inertial mass m_i happens to equal your "gravitational charge" m_g. But Einstein interpreted this equality as an identity: $m_i = m_g$ because gravity simply *is* acceleration. So if you are at rest in a gravitational field—as you will be, if you are sitting on a chair on the surface of the Earth—then the weight you feel is just the effect of the force exerted on your body by the chair and ground in preventing you from falling towards the centre of the Earth. If you were instead in orbit around the Earth, you would be constantly falling because of this same gravitation = acceleration, and since nothing would be preventing you from falling, you would feel no weight.

The equality between the passive gravitational mass m_g and the inertial mass m_i in Newton's theory is sometimes called the Weak Equivalence Principle (WEP), although this is something of a misnomer, since it states the (contingent) *equality* of the two masses, rather than their *equivalence*. Newton tested this equality, and found it to hold within the limits of experimental accuracy he could muster. It has since been tested by Eötvos and others, and is known to hold good to at least one part in 10^{12}. Einstein's principle of equivalence, the Strong Equivalence Principle (abbreviated SEP when contrasted with the WEP, otherwise simply EP), is another matter. Reasoning that there is no way in principle of determining any empirical difference between being at rest in a uniform gravitational field and accelerating uniformly with an equivalent acceleration (or between a body in free fall and the same body moving inertially), Einstein effectively applies Leibniz's Principle of the Identity of Indiscernibles (PII) to assert their identity. This parallels the kind of reasoning he followed in deriving the theory of Special Relativity, described above in Chap. 5. There Einstein took the empirical fact that the velocity of light could not be discriminated from c, no matter what the velocity of its source, to indicate that it is in principle indiscernible from c, so that, by the PII, it is identically c.

Now let's look at some of the startling consequences that Einstein inferred from his EP. The prediction that light would be deflected around a gravitating body is made as follows. According to the EP, the physical effects that could be observed in a system that is in a static, linear homogeneous gravitational field should be indistinguishable from those observable in an otherwise identical system in a reference frame that is in uniform, rectilinear acceleration. Suppose then that the system consists of a light ray travelling in a straight line in empty space from one point to another, say from one side of a cubic container to another: such a straight path in spacetime, the optimal path between two points, is called a *geodesic*. Now turn on a static, homogeneous gravitational field perpendicular to that path. This will have the same effects as if the chamber itself is accelerating for the whole time of travel of the ray. So, because the chamber itself is accelerating, the ray will hit the other wall at a point below where it would have otherwise, and its path, with respect to this accelerating frame, will be curved into a parabola. The EP implies that a light ray in a static, homogeneous gravitational field will also be deflected into a parabolic path in the same way. Similarly, a ray going in a straight line in spacetime around the Sun, which produces a spherically symmetric rather than a linear field, will be bent in towards the Sun (Fig. 7.2).

At this point we can make an illuminating contrast with the Weak Equivalence Principle. For if it is assumed that light consists in massive particles, then these also would get attracted as they passed by a massive body like the Sun. Newton had, in fact, suggested precisely this in Query 1 of his *Opticks*, where he also conjectured

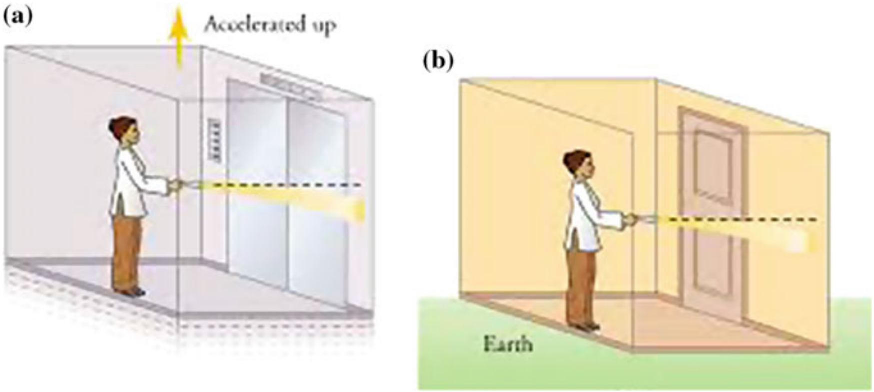


Fig. 7.2 The bending of a light ray by gravity (Illustration from https://archive.cnx.org/contents/e4078a6e-89d4-472a-bb01-8e299d7ba0aa@8)

that a ray of light might be constituted by a stream of (pulsating) particles.[2] His query was taken up in 1801 by the German astronomer Johann Georg von Soldner, who was able to calculate by how much light would be deflected by passing close to the Sun, assuming light to consist of something massive.[3] It actually does not matter what exactly the mass of light is, so long as it is negligible compared to the mass of the Sun—since any body at the same distance from the Sun will experience the same acceleration provided, as the WEP guarantees, $m_i = m_g$. Soldner's calculation yielded a deflection by the Sun of 0.84 s of arc—extremely close to Einstein's own prediction in 1907, as we shall see—although Einstein (along with everybody else at that time) was unaware of Soldner's prediction.[4]

In any case, the point is that if light has mass, then even according to classical physics it will be deflected inwards as it travels around the Sun on its way to Earth. The notion that light has mass, however, would not have been obvious to those supporting a wave theory of light. In this respect it is ironic that 1801—the year Soldner published his prediction—was also the year in which Young made his founding contribution to the establishment of the wave theory, with the result that by the end of the nineteenth century it had become accepted that light consists of waves in the luminiferous aether. Einstein already had reasons for doubting that this was the whole story. One, of course, was his famous relativity paper from 1905 which made the assumption of

---

[2] "*Query* 1. Do not Bodies act upon Light at a distance, and by their action bend its rays; and is not this action (*caeteris paribus*) strongest at the least distance?" (Newton 1730, Bk III, 339) .

[3] In defence of this assumption Soldner wrote: "That light rays have all the absolute [basic] properties of matter one can see from the phenomenon of aberration which is possible only because light rays are truly material. And furthermore, one cannot think of a thing which exists and works on our senses that would not have the property of matter." (Jaki 1978, 948).

[4] As Abraham Pais observes, Soldner's paper "was in fact entirely unknown in the physics community until 1921" (Pais 2005, 200), when it was publicized by the fascist Paul Lénard in an effort to undermine Einstein. For details of this affair and a reproduction of Soldner's paper, see (Jaki 1978).

the aether unnecessary. Another reason was implicit in a second of his seminal papers of that year (his "*annus mirabilis*"), the one proposing the light quantum hypothesis that lay at the basis of quantum theory. Einstein interpreted this as a hypothesis about the way light is emitted and absorbed by matter, however, and was uncertain as to its implications for Maxwell's theory and the propagation of light. Even so, he had every reason to suppose that light is massive. For in a third paper of 1905 he had drawn what was perhaps the most momentous consequence of his Special Theory of Relativity, namely the equivalence of mass and energy encapsulated in his famous formula $E = mc^2$, according to which anything with energy has a mass. Anyway, the upshot is that *if* light has mass then the path of a light ray will be bent as it goes around the Sun, and you can calculate by how much.

According to the Strong Equivalence Principle, however, anything going around the Sun, irrespective of whether it has a mass or what the value of its mass might be, will be in free fall toward the Sun (so that it "feels" no weight), and will thus actually be moving inertially. But an inertial trajectory is by definition a straight line in spacetime. This means that, since the trajectory of light around the Sun is observably bent, so that the stars against which we are viewing the Sun seem to appear farther out from the Sun than they otherwise would, it is the spacetime itself that is bent: the light has a straight trajectory in a warped spacetime.

Einstein was also led to the idea of the warping of spacetime by a second thought experiment involving a rotating disc.[5] For supposing such a disc is rotated at speeds approaching c, the length of its circumference C should be Lorentz-contracted relative to its centre. The length of the radius R, on the other hand, being perpendicular to the direction of motion, will be unaffected. An observer on the disc will therefore be able to fit more lengths equal to R onto the circumference of the disc. So the ratio of C to R will be greater than 2π. The EP thus predicts that an observer in a gravitational field will also find this to be so: space will be non-Euclidean.[6]

Applying the EP to the same example of the rotating disc Einstein inferred the necessity of gravitational time dilation. For suppose we attach a clock A to the disc some distance from the centre, and we are looking down on it from a bird's-eye view, at rest with respect to the centre of the disc. The farther out it is along the radius, the faster it will be rotating, and so the slower the clock will be running relative to a clock B at the centre of the disc, due to (special) relativistic time dilation. Here A's being attached to the disc keeps it at a constant radius from the centre, so it is constantly being accelerated inwards, from our point of view. The farther out it is, the greater will be the acceleration. Now according to the EP the effects produced by this acceleration inwards would be the same as those produced in an equivalent

---

[5] Although Einstein only published these considerations about rotating discs later, they appear to have been instrumental in persuading him of the necessity of a curved spacetime. See Janssen (2014, 192–8) and Dieks (2004) for lucid discussions of these and related points.

[6] There is, admittedly, a slight hitch to this reasoning, as the time coordinates of the observer at rest and the accelerating observer are not the same. The rotating clock does not return to the same point in spacetime after one revolution. It would trace a spiral in spacetime. Nevertheless, the point stands that space would be warped by gravitational fields, according to the EP. See Janssen (2014, 181) for a discussion.

gravitational field acting radially outwards, whose strength is increasing away from the centre. So a clock higher up in such an inverted gravitational field will run slow compared to one nearer the centre: the stronger the field potential, the greater the time dilation. Therefore, in a normal gravitational field which gets stronger towards the centre, clocks will run slower the closer they are to the centre.

In his paper of 1907, Einstein gave a simpler argument for gravitational time dilation based on the EP without reference to rotation (Einstein 1907). Because the effects of a gravitational field are indistinguishable from the effects of being in a frame accelerating towards the centre of the field, light reaching a receiver lower down in the field will have its wavelengths dilated, so that its frequencies will be shifted towards the red end of the spectrum. This is the gravitational red shift alluded to earlier.[7] By the same token, the oscillations of any other device that could serve as a clock will be spaced further apart, yielding a gravitational time dilation.

Four years later Einstein amplified on these predictions and the numerical calculations underlying them. According to the formula he derived on the basis of the EP, a ray of light going past the sun would "undergo deflection to the amount of 0.83 s of arc", and "as the fixed stars in the parts of the sky near the Sun are visible during total eclipses of the Sun, this consequence of the theory may be compared with experience" (Einstein 1911, 108). Slightly improved values for the terms in Einstein's formula yield a deflection of about 0.87 s, but this value is based only on the EP, rather than on the field equations of the General Theory, which he had not yet derived in 1911. But he had done so by 1916, and to his great satisfaction he was able to calculate from them a value for the deflection of twice the value predicted with the EP alone, namely 1.75 s of arc.[8] This prediction eventually led (after the intervening First World War) to Eddington's famous expedition to the Island of Principe off the coast of Africa to observe a solar eclipse, and to his dramatic endorsement of Einstein's (revised) prediction in 1919 (Dyson et al. 1920). Einstein's second prediction in the 1911 paper, that of gravitational time dilation, came out the same in GR, and has been verified experimentally to an accuracy of about 1%. Since his reasoning there involves only some relatively straightforward mathematics, it is worth discussion for the light it throws on our topic of the flow of time.

Suppose a light impulse is emitted from a source S_2 in a uniformly accelerated system K' towards a detector S_1 with a frequency ν_0 relative to a clock U in K', with an acceleration g. Now suppose an unaccelerated system of reference K_0, relative to which K' has a velocity of 0 when the light is emitted. If the distance the accelerated

---

[7] Actually, as we'll see, there are several other components to the overall red shift of the light from distant galaxies. One is due to the fact that each of the galaxies is moving apart from all the others with the overall expansion of space: this is the famous Hubble red shift, our best indicator of the expansion of the universe. The latter is due to the expansion of space itself over cosmic distances, and so comes into play when assessing the redshift of light from distant galaxies. But it is a very small effect over distances such as that of the Earth from the Sun.

[8] The discrepancy is due to the fact that for the original prediction Einstein applied the EP in a flat background space. As we shall see, the EP properly only applies locally, in arbitrarily small regions. The curved metric of GR knits these local regions together in such a way as to double the amount of the curvature.

system travels in K_0 is h, its velocity at the time of arrival will be $v = gt$ with $t = h/c$ in that frame, giving $v = gh/c$. Then when the light reaches the detector S_1 its frequency "relatively to an identical clock in S_1" is given by the Doppler formula:

$$\nu_1 = \nu_0(1 + v/c) = \nu_0\left(1 + gh/c^2\right) \tag{7.3}$$

But by the Equivalence Principle, this is equivalent to the situation where the same light impulse is emitted from S_2 to S_1 in a uniform gravitational field, where gh is the gravitational potential Φ. Suppose the light is emitted from the surface of the Sun, and received at the Earth. The (negative) gravitational potential Φ of the Earth in the Sun's field is $g_S h$, so the Doppler formula gives

$$\nu_1 = \nu_0(1 + v/c) = \nu_0\left(1 + g_S h/c^2\right) = \nu_0(1 + \Phi/c^2) \tag{7.4}$$

On re-arranging, we get

$$(\nu_0 - \nu_1)/\nu_0 = -\Phi/c^2 \tag{7.5}$$

corresponding to a shift of the spectral lines towards the red which Einstein computes to be 2×10^{-6}, a prediction later verified by the experiments of Pound and Rebka (1960) and Pound and Snider (1965).[9]

Here we have assumed (as Einstein implicitly did) that the Sun's gravitational force is very much greater than the Earth's ($g_S \gg g_E$). If we were just considering light emitted in the Earth's gravitational field—for instance, from a space shuttle in orbit around the Earth at a height h above the surface of the Earth—then the light from the shuttle would be blue-shifted. But for light from the Sun detected at the Earth's surface, this blue shift would be correspondingly small compared with the red shift resulting from the Earth's being in the Sun's field.

Regarding his Eq. (7.4), Einstein remarked that superficially it "seems to assert an absurdity" (Einstein 1911, 105). "If there is constant transmission of light from S_2 to S_1, how can any other number of periods per second arrive in S_1 than is emitted in S_2?" (105–6). But ν_0 is the frequency measured by the clock U in S_2, "while ν_1 denotes the number of periods per second with respect to the identical clock in S_1. Nothing compels us to assume that the clocks U in different gravitational potentials must be regarded as going at the same rate" (106). In other words, such an objection rides on the same fallacious reasoning adopted by those rejecting special relativity (Chap. 6): to claim that the frequency has changed is to assume that time flows at the same rate higher up in the gravitational potential as it does lower down, that there is an absolute time that flows everywhere at the same rate. In fact, consistency requires

---

[9] It may appear odd that in the above derivation Einstein uses the classical Doppler formula and not the relativistic Doppler formula that he had derived in his original SR paper (Einstein [1905] 1923, 56): $\nu_1 = \nu_0\,\gamma(1 - v/c)$. But as Einstein had already shown in his 1907 paper, by expanding γ in a power series, the formula can be shown to be correct "to a first approximation" in which terms of order v^2/c^2 are neglected.

that “if we measure time in S_1 with the clock U, then we must measure time in S_2 with a clock which goes $(1 + \Phi/c^2)$ times more slowly than U when compared with it at one and the same place” (106). Since this would be true of any process at all that could serve as a clock, we have to say that *time itself flows at (relatively) different rates at different heights in the gravitational field*.

This phenomenon of gravitational time dilation has particularly interesting implications for the behaviour of matter near black holes. A black hole is formed when gravitating matter reaches such a high density that its acceleration inwards becomes unstoppable. This is a prediction even in the Newtonian theory of gravity, where matter’s mutual attraction causes the gravitational collapse, whereas in GR it is due to the extreme warping of spacetime. According to theory, the collapse continues indefinitely, leading to a singularity in spacetime where GR would no longer apply and the density of matter would be infinite. Classically it was predicted that at a certain radius from the centre of the collapsing matter, the escape velocity—the velocity needed to break out of the gravitational field—is greater than c, so not even light can escape. This is also a consequence of Einstein’s theory, as was pointed out to him by Schwarzschild in late 1915. Consequently, the limiting distance from the singularity at which this occurs is known as the “Schwarzschild radius”: the warping of spacetime is so extreme that nothing falling into the hole can escape once it is within this radius.[10] At this distance the time dilation is so great that all processes occurring in any object falling into the hole come to a standstill—relative to a standpoint outside the Schwarzschild radius. From the perspective of the falling object itself, time—its proper time—will pass normally, but its passage will slow to a stop relative to the time coordinate of a reference frame of any observer situated “higher up” in the gravitational field outside the Schwarzschild radius. This is another illustration of the fact that time flows at (relatively) different rates at different heights in a gravitational field (leading to different elapsed proper times along different paths between two points in spacetime), even while the time elapsing for a massive system (such as a clock) along its own particular path is invariant. It is also serves as another counterexample to the classical conception of becoming as taking place in a steady progression of instantaneous states from the past towards the future. Although all clocks record time lapsing locally, the warping of time near black holes is such that no clock external to the Schwarzschild radius will be able to chronicle the passage of time undergone by anything falling into the hole.[11]

---

[10] Here we should make a couple of caveats: first, as Stephen Hawking showed, because of quantum effects, black holes can in fact evaporate by (very slowly) leaking radiation; second, as has been discovered very recently by Joseph Polchinski and colleagues (Almheiri et al. 2013), quantum effects should result in the production of a seething maelstrom of particles at the event horizon, whereas the Equivalence Principle predicts that the free fall of a body through the event horizon should be an inertial motion. This clash of predictions is known as the Firewall Paradox, and may not be resolved until we are closer to a theory of Quantum Gravity that supersedes GR and Quantum Theory while accounting for their successes.

[11] Richard Muller argues on this basis that “actually, there are no black holes” (Muller 2016, 86), reasoning that “it takes infinite time to form a black hole, measured in our time coordinate” (87). Meanwhile, the fall takes only ten minutes in the rest frame of matter that is accelerating towards

The notion of a curved spacetime is admittedly difficult to grasp, and it is easy to get confused about what is curved according to GR, and what is not. A basic concept is that of a *geodesic*, which can be understood by comparison with the classical definition of a straight line as the shortest distance between two points. Imagine, for example, that we are concerned with the shortest distance between two points on the surface of a sphere. This is given by a great circle (an arc on the surface whose radius is that of the sphere) joining the points: any other curve joining them will be longer. (This explains why, when flying from, say, Toronto to Paris, the plane traces an apparently northerly route, curving up through Labrador and over Greenland, rather than following what looks like a straight line on the usual map, namely, Mercator's projection. The route is actually the shortest distance: the curve bending north is an artefact of the mapping, which projects the surface of the sphere onto a plane, bending the straight line into a curve and distorting the apparent sizes of northern and southern land masses in the process.)

Analogously, a geodesic in spacetime is an extremal path between two spacetime points. In Minkowski spacetime, as we saw above, a straight line is the *longest* spacetime interval between two points in timelike separation. It represents the path of something moving inertially from one to the other, and along such a path the proper time is a maximum: any other path between the points represents something accelerating at some point in its motion, and it is necessarily shorter. But with the Equivalence Principle the concept of inertial motion undergoes a radical change. Now a body that is falling freely in a gravitational field is in fact moving inertially. The Moon, for example, is in constant free fall towards the Earth. This means that (neglecting the gravity of the Sun and other planets) it is moving inertially, and thus moving in a straight line in spacetime. It does indeed move in a curved path around the Earth, but that is a projection of the curved 4-D spacetime due to Earth's gravity onto a 3-D space. (It is the same with light rays bending around the Sun.) This is analogous to how the projection of the surface of a three dimensional sphere onto a 2-D map (using Mercator's projection) distorts straight lines on that surface (the shortest distance from Toronto to Paris) into curves on the map.

A second example, due to Richard Feynman, takes us back to Bacon's clocks ticking at different rates in a gravitational field. A clock on the Earth's surface is not performing inertial motion: it is rotating with the Earth, and also, according to the EP, it is accelerating away from the Earth's centre. So it takes less time to go from, say, 12:00 noon to 12:01 than some second clock travelling on an inertial trajectory, and thus taking maximum time. Here Feynman presents a poser: how would we get the second clock to record such a maximum time? The answer is: throw it straight upwards with sufficient momentum that it returns to Earth at exactly 12:01 according to the clock that had stayed on the Earth's surface! For that trajectory will be inertial, tracing a straight line in spacetime. The thrown clock will then record more than

---

the centre of the black hole, while from the standpoint of this frame "at eleven minutes, the time outside has gone to infinity and beyond" (83). Granted; but Muller fails to acknowledge that this entails that there is no such thing as the "universe at an instant of time" which he presupposes in his concept of *Now* as "the leading edge of time" (293), or of which he supposes one could have a "God-like complete knowledge" (121).

a minute has passed (assuming it is sufficiently accurate). This, we may note, is consistent with the idea that clocks higher up in the Earth's gravitational field will run faster: the thrown clock gradually gains on the one that remains on Earth until it returns. (When it returns to the Earth's surface, however, it will be ticking at the same rate as the one that remained on Earth.)

So the upshot of this discussion is that in the world of general relativity, there is no global perspective from which time passes at a constant rate.[12] Just as in the special theory, clocks that are moving relative to the observer run slow, so in the general theory clocks closer to the source of a gravitational field also run slow. In both theories, the time elapsed between two spacetime points is maximal along an inertial path joining them, and there are different rates of flow along different paths through spacetime.[13] This conclusion will be bolstered by the lines of argument pursued in Sects. 7.4 and 7.5.

7.3 Interpreting General Relativity

Let us now take a look at the main features of the theory of general relativity. Even though the devil is in the (highly technical) details, I will restrict the formal apparatus to a minimum. The theory was recognized as a major accomplishment ever since its triumphant endorsement by Eddington. But from its inception it has been something of a tangle of philosophical and mathematical components, and its reception by leading philosophers of physics was heavily influenced by their own philosophical proclivities.[14] Leaving the interpretations of his peers to one side, this was no less true of Einstein himself. His interpretation of the theory even in creating it was

---

[12] Lee Smolin has argued that a global time function is definable in Shape Dynamics, in which only the local shapes of Riemannian 3-geometries are dynamical (see Barbour 2011 for an introduction). This allows the slicing of spacetime into three-dimensional spaces of constant mean curvature, the so-called CMC foliation, where the spaces evolve in a "dynamically determined preferred global time" (Unger and Smolin 2015, 420). But as Smolin concedes, "the effects of the preferred global time, if it exists, are not detectible in experiments at less than the scale of the whole universe" (491–92). Such a time tracks the cosmological expansion of space, but not the local becoming of events in it.

[13] Cf. Roberto Torretti: "Gravity does not therefore impair the time-keeping functioning of natural clocks; it shapes the several strands of time measured by differently placed clocks" (Torretti 1999, 291). Cf. also George Ellis: "The relative flow of time along different world lines may be different: that is the phenomenon of time dilation, caused by the varying gravitational potentials represented by the metric tensor [78]. But this does not mean it is not well defined along each world line." (Ellis 2012, 12). All this should be compared with Roberto Mangabeira Unger, who claims that "None of the classical empirical tests of general relativity … have any direct or proximate relation to time" and writes dismissively of "so-called time dilation" (Unger and Smolin 2015, 191).

[14] That is true of the conventionalist reading of GR given by Henri Poincaré and Moritz Schlick, of the attempts by neo-Kantians like Ernst Cassirer and Hans Reichenbach to assimilate it to a revised neo-Kantianism, and of the objective idealism that Hermann Weyl and Arthur Eddington saw to be its main lesson. See Ryckman (2014) for a highly informative account of these philosophical interpretations of general relativity.

heavily coloured by his own philosophical predilections, especially concerning the interpretation of coordinate systems and the origin of inertia, where he took inspiration from Mach's philosophy. Indeed in 1918 he saw the theory initially as a grand confirmation of what he called "Mach's Principle", only to retract that assessment later. Tangled up in this too are certain misconceptions about what general relativity actually consists in—whether, for instance, it means that all motions are relative, or that all reference frames are equally valid for describing the phenomena, or that the equations of the theory must be put in a covariant form, free of reference to any particular coordinate systems. So I will do my best to untangle some of this confusion, and to try to lay bare the essentials of the theory as it appears to us now, after a century of work by historians and philosophers of science.

As a first step, we need to see what prompted Einstein to create GR. Why, after developing his Special Theory of Relativity and its revolutionary implications, such as the equivalence of mass and energy, did Einstein seek to improve upon it and replace it with a more general theory? First, if we think of GR as a theory of gravity, then it was perfectly clear to those who accepted the Special Theory that the Newtonian theory of gravity, successful as it was, needed emendation in the light of SR. For according to Newton, two masses are supposed to attract each other mutually at every instant. Such an instantaneous action at a distance had been heavily criticized by Leibniz and by Mach as being incompatible with the principles of mechanics. Now this criticism seemed to be confirmed by the Special Theory, which showed that no action or influence could propagate faster than the speed of light, c. So the Newtonian theory was obviously in need of correction.

Of course, the very naming of SR as "special" indicates that Einstein regarded it as not sufficiently general. The most salient respect in which that is so is the special status in the theory of inertial frames. Why should one class of reference frames have any claim to priority over any other? This seemed to Einstein as arbitrary and unfounded as the status of aether in the classical theory of electromagnetism. Perhaps, then, in a properly general theory, inertia would be explained away, or at least explained to result from our having taken a perspective that is not appropriately general. As we have seen, Einstein had his reasons for thinking that the explanation of inertia was deeply bound up with an acceptable theory of gravity, as suggested by his Equivalence Principle. But once this principle was put to work, it implied, as we saw above, both that spacetime itself must be curved and that time would be dilated in gravitational fields, both of which results implied that the Special Theory would need extending, since it was based on Minkowski's flat spacetime:

> The introduction of coordinate systems accelerated relatively to each other as equally legitimate descriptions, such as they appear conditioned by the identity of inertia and weight, leads, in conjunction with the results of the special theory of relativity, to the conclusion that the laws governing the arrangement of solid bodies in space, when gravitational fields are present, do not correspond to the laws of Euclidean geometry. An analogous result follows for the motion of clocks. (Einstein 1921, 247–48)

But for Einstein himself the most compelling motivation for a generalization of his special theory was his philosophical commitment to the relativity of motion, and the aim to extend it to all reference systems, not just inertial ones. As we have seen, he followed Huygens and Mach in insisting on the relative nature of all motion, and interpreted this as a relativity to reference frames conceived as co-ordinate systems defined by material rods and clocks. It should therefore encompass a relativity not just to unaccelerated coordinate systems, as Einstein understood the Special Theory to deliver, but also to accelerated coordinate systems.[15] On this reading, given the Equivalence Principle, the Special Theory is deficient in not applying to gravitational phenomena[16]; whereas, Einstein believed, a properly general theory would render all motion relative, so that there would be no privileged class of coordinate systems. In keeping with such a perspective, Einstein made three distinct attempts to articulate constraints on what form the General Theory should take in order to comply with the relativity of motion in general, as noted in the introduction to this chapter: compliance with the EP; covariance of the field equations; and compliance with Mach's Principle. As noted, none of these delivered the required universal relativity of motion.

The importance of this for our topic is as follows. If all motion is relative in the general sense that Einstein desired to capture, then time is simply relative to the coordinate system adopted, whether inertial or not. Einstein himself implied on occasion that this would make time simply a facet of our representation, even if not entirely subjective. Such a view has been instrumental in persuading physicists and philosophers alike that time has no fundamental status in spacetime theories, and has fed into the conceit that time is altogether dispensable in fundamental physics. I have already argued against this view in the context of the Special Theory, but if one regards that theory as a provisional step towards a properly relativistic theory, as Einstein did initially, then the case needs to be made that this is not a viable interpretation of GR. I should add that it is no disrespect to Einstein to say that we now understand these matters differently than he did as he set about creating GR. Such retrospective reformulation of theoretical discoveries is a normal and underappreciated part of the process of scientific advance. And in this case, our current understanding is largely due to Einstein himself, who modified his views in dialogical exchanges with his peers as they reacted to his ground-breaking ideas.

Before discussing how GR falls short of a universal relativity of motion, though, we should first lay out in schematic form the main features of the theory.

First, Einstein had tried to develop the field equations using the EP and the consequence that he had deduced from it in 1911 for homogeneous gravitational fields, that the speed of light c is variable, generalizing this to apply to arbitrary static fields. As a generalization of Newton's gravitational theory, expressed using the Poisson equation as $\nabla^2\Phi = 4\pi\sigma$, with Φ the gravitational potential and σ the mass density,

---

[15] In a 1907 yearbook article on relativity Einstein began: "So far we have applied the principle of relativity … only to non-accelerated systems. Is it conceivable that the principle of relativity should apply to systems that are accelerated with respect to each other?" (Isaacson 2007, 148). Similarly, in 1919 he asks rhetorically: "Should the independence of physical laws of the state of motion of the coordinate system be restricted to the uniform translation of coordinate systems in respect to each other?" (Einstein 1919; in 1954, 230)

[16] "The special theory, on which the general theory rests, applies to all physical phenomena with the exception of gravitation; the general theory provides the law of gravitation and its relations to the other forces of nature." (Einstein 1919; in 1954, 228–29)

he proposed the equation $\nabla^2 c = kc\sigma$, where k is a constant and c is the variable speed of light. As he came to see, however, the resulting field equation violated the conservation of energy and momentum, unless he subtracted a certain quantity from the left-hand side of the equation. But this modification required the limiting of the EP to arbitrarily small regions of spacetime—a surprising consequence, given that the fields in question were homogeneous. Other difficulties led him to abandon this approach. Nevertheless, the lesson was that the EP had to apply only locally: the Special Theory is valid only in "the important limiting case of a constant gravitational potential", with the result that the EP applies only to "infinitesimal spaces" (Einstein 1912, 1063; quoted from Torretti 1996, 137).

As we have seen, Einstein had also inferred the curvature of spacetime predicted by the EP by means of his ingenious thought experiments involving rotating coordinate systems, which entailed that space itself could not be Euclidean. Such a curving of spacetime implies that the geodesics—the paths followed by non-rotating test particles—are not straight lines in a flat, Minkowskian spacetime. Still, they must reduce to them in the limit. According to the (local version of the) EP, the metric of general relativistic spacetime, $\mathbf{g}$, must agree with the Minkowski metric η in a small neighbourhood of any spacetime point. This is analogous to the way in which, according to Gauss's theory of surfaces, the spatial metric of the surface of a sphere must agree locally with those of the almost flat neighbourhoods of all the points on its surface, an analogy that occurred to Einstein in late 1912. What this analogy suggests is the piecing together of a non-uniform gravitational field from all the local charts covering its near-uniform small neighbourhoods. (We call these local charts conforming to SR "Lorentz charts", making up an "atlas".) So the invariant spacetime interval (see Chap. 5 above) given in differential form for flat Minkowski spacetime by

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \tag{7.6}$$

needed to be generalized to allow the four spacetime coordinates, t, x, y, and z, to be varied arbitrarily, provided the values of components of the metric field $\boldsymbol{g}$, denoted $g_{\mu\nu}$, are adjusted appropriately. This requires sixteen terms for the spacetime interval:

$$ds^2 = g_{00}dx_0^2 + g_{01}dx_0dx_1 + \cdots + g_{12}dx_1dx_2 + \cdots + g_{33}dx_3^2 \tag{7.7}$$

These values of $g_{\mu\nu}$, may be set out in a square array as follows:

$$\begin{aligned} g_{\mu\nu} = {} & g_{00}g_{01}g_{02}g_{03} \\ & g_{10}g_{11}g_{12}g_{13} \\ & g_{20}g_{21}g_{22}g_{23} \\ & g_{30}g_{31}g_{32}g_{33} \end{aligned}$$

This is known as a tensor; it is symmetric, so that $g_{12} = g_{21}$ etc., and there are therefore only ten independent terms. Now the coordinates are completely general. As we have seen, Einstein identified coordinate systems with reference frames; consequently, because $\boldsymbol{g}$ encodes all possible linear coordinate transformations, it suggested to him the equivalence of all frames of reference that he believed would guarantee the universal relativity of motion.

Unfamiliar with the mathematics of tensors, Einstein sought help from his friend Marcel Grossman, who introduced him to Riemann's and Christoffel's generalization of Gauss's work, and to the absolute differential calculus for tensor fields of Ricci and Levi Civita. The key here is that the curvature of a surface can be characterized internally—without reference to any supposed container space in which it might be thought to be embedded—in terms of the tangent vectors at each point. In accordance with the requirement that $\boldsymbol{g}$ agree with η locally at each point, curves in the metric $\boldsymbol{g}$ may then be distinguished as *timelike*, *spacelike* or *null*, by reference to their tangent vectors at each of these points.

Since the metric field determines which motions are inertial, it follows that a chargeless, non-rotating test particle of non-zero mass, subject to no external non-gravitational influences, must describe a timelike geodesic (while photons will describe null geodesics). This requirement is known as the *Geodesic Principle*, and was regarded by Einstein as the core of the SEP. Initially introduced by him as a postulate, Einstein and Grossman (1913) derived it as a theorem under the assumption that matter is a pressureless dust, and later Einstein derived it from his field equations with the aid of some younger collaborators in a (somewhat controversial) proof. It is important to understand, as Torretti has stressed, that this principle does not automatically follow from the fact that the metric is locally Minkowskian. Even though a freely falling particle π will trace a geodesic in the Minkowski metric for that local domain to the extent that the field is homogeneous there, these local domains are not restrictions of some global Minkowski metric "any more than the Euclidean metrics which show up, say, in the street plans of Mannheim and Manhattan are the restrictions to these boroughs of a Euclidean metric defined on the entire Earth" (Torretti 1996, 151). But if the global metric $\boldsymbol{g}$ is approximated by the local flat metric η on a small neighbourhood of each worldpoint, has the Lorentz signature, and η agrees with $\boldsymbol{g}$ at the origin of the pertinent local Lorentz chart, then the Geodesic Principle consists in the stipulation that "if $\int d\tau$ is now made to stand for the length of π's worldline as determined by $\boldsymbol{g}$—i.e. for the proper time measured along it by a natural clock at π—then the variational principle $\delta \int d\tau = 0$ is obeyed by the freely falling particle between any two events in its history" (151).

In Einstein's theory the geodesic principle follows from the metric connection, which threads together the various tangent spaces to a timelike spacetime path in such a way that the path has the same chronometric significance as in SR. That is, there is no contribution to the proper time elapsed along any timelike path beyond that contributed by the local Lorentz frames.[17] This contrasts, for example, with Hermann Weyl's celebrated attempt at a unified theory of gravitational and electromagnetic forces (Weyl 1918), according to which the rate at which a clock ticks will depend on its history. Einstein objected that according to Weyl's theory a clock encountering a varying electromagnetic potential in a static gravitational field would return to its starting point ticking *at a different rate* than one that had remained at the starting point, contrary to all empirical evidence. This is different from the case in the Twin Paradox, where the clocks will be running at the same rate when they reconvene, despite recording differing times having elapsed for their journeys. But Weyl's theory would imply that identical particles taking different paths through spacetime would reconvene with time running at different rates, and therefore with different natural frequencies and different rest-masses. Roger Penrose has stressed this point: "in Weyl's geometry, not just clock rates but also a particle's *mass* will depend upon its history" (Penrose 2005, 453). Such a consequence is precluded in Einstein's theory by the metric connection embodied in the geodesic principle, which preserves the chronometric significance of motions along geodesics embodied in the inertial motions of classical and special relativistic theory. This is a vital and relatively unappreciated point about time in general relativity, and we will return to in Chap. 8.[18]

Now the key to Einstein's derivation of his field equations is the idea that the metric field $\boldsymbol{g}$ describes the geometry of spacetime induced by the presence of the distribution of mass-energy in a given region. Einstein took his cue about how to represent the latter from Max von Laue, who had contended that in each domain of physics the dynamical state of a material continuum should be represented by a tensor built up from quantities equivalent to the classical mass-energy density, the momentum density, and the components of the classical stress tensor. This is the stress-energy tensor, a symmetric tensor of rank 2, which we designate by $\boldsymbol{T}$.[19] The ten equations determining the metric field would then have the form

$$\Gamma_{\mu\nu} = \kappa \mathrm{T}_{\mu\nu} \tag{7.8}$$

where κ is a constant, and $\Gamma_{\mu\nu}$ is the as yet undetermined gravitation tensor, to be constructed by differential operations out of the metric tensor $g_{\mu\nu}$.

One of the great advantages of this formulation in terms of tensors is that the equations of the theory are independent of the coordinate system used to expressed them.

---

[17] This feature of relativity theory is usually called the "clock hypothesis". As I have argued in my (2010), however, it is not a separate hypothesis: in SR it is a consequence of the definition of proper time, and in GR it follows from the Equivalence Principle.

[18] See also the lucid discussion in Penrose's (2016, 52–59), where he details the later fortunes of Weyl's theory as the foundation of the so-called "gauge theories" of strong and weak interactions in the standard model of particle physics.

[19] A point to note in passing here is that, despite the great success of his using the energy-momentum tensor in formulating GR, later in life Einstein became quite sceptical about representing matter by such tensors. As Torretti reports (Torretti 1999, 295–6), he came to regard them as idealized models rather than fundamental physical entities, "as purely temporary and more or less phenomenological devices for representing the structure of matter" (Einstein and Infeld 1938, 209).

Thus they satisfy the requirement of *general covariance*, namely, that the equations of the theory should retain the same form under any continuous and differentiable transformations of the coordinates. This was recognized immediately by Einstein as an advantage, although for reasons I will not pursue here, he was led by some plausible physical arguments to think that this requirement could not be met by the equations of gravity, and his and Grossman's theory (Einstein and Grossman 1913), the so-called *Entwurf* theory, held only for an ample but restricted class of coordinate systems. This precipitated Einstein's three-year odyssey from the *Entwurf* theory to his presentation of the generally covariant field equations in November 1915, involving his famous "hole argument"—a story that has been told eloquently elsewhere.[20] Finally, in 1916 he derived the Einstein field equations (EFE) that are the fulfilment of his quest for a general relativistic theory of gravity;

$$R_{\mu\nu} - {}^{1}\!/_{2} R g_{\mu\nu} = (8\pi G/c^4) T_{\mu\nu} \tag{7.9}$$

where $\boldsymbol{R}_{\mu\nu}$ is the Ricci curvature tensor, R is the scalar curvature, and $g_{\mu\nu}$ and $T_{\mu\nu}$ as above. On the left are terms representing the metric $\boldsymbol{g}$ and its first and second derivatives, and these are equated with the stress-energy tensor $\boldsymbol{T}$, representing all the sources of gravity save for the gravitational energy itself. The stress-energy tensor, however, is very much an idealization, for, as Torretti explains, "we are quite incapable of accurately representing the distribution of matter and energy in the universe and, if we could do it, we would be unable to solve the EFE for a matter-energy tensor field of the requisite complexity" (Torretti 1999, 77). This is because the EFE are a non-linear set of equations: the structure of spacetime depends on the mass-energy distribution, which in turn affects the structure of spacetime. In practice, as we shall see below, the field equations can only be solved for highly idealized cases where they can be reduced to a set of linear equations.

The first such solution was provided within two months of Einstein's publication of the EFE (6.9) by the German physicist Karl Schwartzschild (1873–1916), who used them to model the motion of a small body (like Mercury) in orbit around the Sun. The idealizations Schwartzschild assumed were that the gravitational field is static in time and spatially symmetric in space, and that it converges to the flat Minkowski metric at infinity. Mercury is modelled by a nonrotating object of negligible mass, and the influence of all other masses in the solar system is ignored. Using this model Schwartzschild was able to give a precise confirmation of the predictions Einstein had made (using approximations) of the advance of the perihelion of Mercury by 43″ per century. In Sect. 7.4 we'll return to Einstein's objection to one of Schwartzschild's assumptions, and consider his own and other exact solutions.

Now let us turn to the issue of Einstein's quest to establish a universal relativity of motion, beginning with the Equivalence Principle. Does this principle not embody the idea of the universal relativity of motion, given that acceleration is relative to the reference frame chosen? Actually, it does not. To see this, imagine an astronaut in a

---

[20] For a comprehensive account, see (Norton 2018); see also (Torretti 2000), and for concise accounts (Torretti 1999) and (Janssen 2014).

space shuttle orbiting the Earth. According to the EP, since she is in free fall towards the Earth, everything will seem to her exactly as if she were deep in outer space, far from the gravitational field of any planet. She will be moving inertially. If we describe this from the perspective of Ground Control on Earth, however, she will be orbiting around it in the shuttle, and, like the Moon, constantly accelerating towards the Earth. So, it may seem, acceleration is relative to whichever frame of reference we have adopted. But this is to ignore the fact that the presence of the Earth's gravitational field is not transformed away by such a change of frames. The astronaut is moving inertially in a straight line in a spacetime warped by the gravity of the Earth, and, in a reference frame co-moving with her, Ground Control is accelerating away from her because of the equivalence of gravity and acceleration; in the reference frame in which Ground Control is at rest, the astronaut's shuttle is accelerating towards the Earth in order to stay in orbit. But that is an effect of having adopted an accelerating frame of reference; it does not alter the fact that the astronaut is really moving inertially and that the gravity of the Earth has warped the spacetime around it. It does not transform away the inertial motion, but simply redescribes it. So this is unlike the case of the relativity of inertial motion in SR, where by adopting a different inertial frame we can transform away an inertial motion with a given velocity. In GR, we do not transform away inertial motion by adopting an accelerated frame, we just give an equivalent description of the same scenario.[21]

Similar objections apply to Einstein's second attempt to guarantee a universal relativity of motion. Once he had arrived at his (fully) generally covariant field equations in 1915, he supposed that this covariance, in allowing arbitrary coordinates and coordinate transformations, would automatically extend the principle of relativity from uniform motions to arbitrary motions. This was because he persisted in thinking of general relativity as being constituted by the equivalence of all coordinate systems, accelerated as well as unaccelerated (relatively to one another), for providing legitimate descriptions of the phenomena. But as we have just seen, the equivalence provided by covariance is between descriptions of the same scenario in different coordinate systems, but not a transformation between physically equivalent states of motion. This was first brought to Einstein's attention by Erich Kretschmann in 1917, who pointed out that even a theory like Newton's, which privileged certain motions as absolute, could be put into a covariant form. As he explained, the Lorentz transformations in SR capture a symmetry of Minkowski spacetime, mapping the set of all geodesics representing inertial paths back onto itself, thereby representing the relativity of uniform motion. But the set of all geodesics of the various spacetimes of GR, corresponding to different matter distributions, does not have any such non-trivial symmetries, and is therefore not associated with a principle of relativity of motion. There is no privileged class of coordinate systems, so that in this sense all time coordinates are on a par; on the other hand, though, according to the geodesic principle, it is along the geodesics that proper time is measured, and these are not rendered non-inertial by any coordinate transformation.

---

[21] Cf. Torretti (1996, 135), to whose explanations I am indebted here.

Einstein's third attempt at capturing the gist of general relativity was inspired by his reading of Mach, as noted above. This attempt derived from his conviction that if all motion is relative to other ponderable bodies, then there could be no such thing as motion relative to empty space. He therefore followed Mach in rejecting Newton's argument for absolute motion. Newton had argued that a sure criterion of its being the Earth that is really rotating in absolute space is the centrifugal force experienced on Earth's surface, a force absent in a non-rotating body. Mach had asked rhetorically whether an equivalent force might not also be experienced if the universe were rotated around the Earth. Since Einstein held space to be determined by the distribution of mass-energy, he took Mach to be making the positive conjecture that a body's inertia is a consequence of the distribution of all the other matter in the universe, dubbing this "Mach's Principle". This issue needs setting in the context of the whole question of the relativity of rotational motion, where there are other implications to be explored for the understanding of time.

7.4 Rotational Motion, Inertia and Mach's Principle

This strand of Einstein's thinking about relativity has its origins in the classical debate about the nature of space. As will be recalled from our discussion in Chap. 5, Mach's empiricism aligned him with Huygens's stand on the relativity of motion, that the only intelligible motion is motion with respect to other bodies. Leibniz had raised this to a general principle, the Equivalence of Hypotheses: equivalent descriptions of the phenomena can be had whatever hypothesis is made about which bodies are at rest. Leibniz qualified this relativity, however, as applying to motions only insofar as they are regarded as *geometrical*. In principle one can identify which of two bodies in relative motion is really moving by reference to the *causes of motion*. In some cases, such as a stagecoach moving along a road, this is straightforward: it is the horses pulling the coach that are expending the effort to produce the relative motion: they are the cause of the motion, which has nothing to do with the road. In others, such as the motion of the Earth, it requires consideration of the "most intelligible hypothesis", which is then to be regarded as (defeasibly) true. Thus, Leibniz argued, the Copernican hypothesis is the most intelligible hypothesis in astronomy, and therefore to be regarded as true.[22]

Nevertheless, Leibniz maintained (no doubt as a result of his apprenticeship with Huygens in the 1670s), that the received laws of collision and inertia constituting the physics of his time were frameable entirely in terms of the relative motions of

---

[22] Leibniz first articulates this view in an unpublished manuscript of 1676 (see Arthur 2013). There he notes the greater simplicity of the Copernican hypothesis in its dispensing with the imaginary epicycles and eccentric circles of the Ptolemaic system, the potential changes in the apparent diameters of the fixed stars and changes of situation of the Earth relative to the fixed stars, and observations of oscillations of hanging lamps, and of tides "impinging only on eastern and western shores" (A VI 3, 105). These things, he concludes, "can be explained more distinctly by the supposed motion of the Earth and its being reduced to a simple cause" (111).

bodies. Consequently, as early as 1677 he had already taken the relativity of motion to preclude the existence of any one preferred space, "absolute space". He did this in manuscripts he did not publish, but composed a full ten years before Newton himself had published anything about space. In one he writes:

> Absolute space is no more a thing than time is, even though it is pleasing to the imagination; indeed it can be demonstrated that such entities are not things, but merely relations of the mind trying to refer everything to intelligible hypotheses—that is, to uniform motions and immobile places—and to values deduced on this basis. ("Motion is not Something Absolute"; A VI 4, 1638; Leibniz 2001, 333)

When Newton did publish in 1687, however, he proposed a thought experiment designed to show the complete untenability of a purely relative motion. This was his famous "bucket thought experiment". Newton asked the reader to consider a pail half filled with water, suspended from a fixed point with a cord (today, a bungee cord works very nicely). The bucket is twisted through many revolutions, and then released so that it spins rapidly about its centre. When the water is at rest, its surface will be flat; when, after several seconds, it has picked up the motion of the bucket, the surface of the water takes on a concave shape, and climbs the side of the pail (actually, in beautiful spiral arms, like a spiral galaxy!). This is due to the centrifugal force, the endeavour of the water to continue moving in a straight line owing to its inertia, which, because of the constraint of the sides of the bucket, tends to move it away from the centre and climb the inclined sides of the pail. Newton's point is that the centrifugal endeavour is a criterion for the reality of this motion. The motion of the water cannot simply be a motion relative to the sides of the bucket with which it is in contact, as Descartes had insisted. For when the bucket is initially set spinning and the water has not yet picked up this motion, there is a large relative motion between the bucket and the water: but the fact that the water has not yet started to rotate can be seen from its flat surface. And when it has picked up the motion, its relative motion to the bucket is nil, but the concave surface is a sure sign that it is now rotating. So its motion is absolute, not relative to the bucket. And an absolute motion, so Newton argued, presupposes an absolute space in which it may be represented.[23]

But Huygens and Leibniz were unconvinced, since for them motion was essentially motion with respect to other bodies. A motion with respect to an unobservable entity like absolute space was unintelligible in principle. Huygens (1690) proposed a mechanical theory of gravity as a kind of deficit of centrifugal force. His idea was that in a vortex of fluid matter, such as the vortex of "subtle matter" he supposed to be caused by the spinning of the Earth, all the particles composing the vortex would have a centrifugal force impelling them away from the centre of the Earth. He hypothesized that any ponderous body had comparatively less centrifugal force than subtle matter occupying the same volume, so that it would suffer a differential force toward the centre—gravity. (Think of the twigs and clumps of flotsam that congregate towards the centre of an eddy in a river.) Huygens also thought he could account for the invariance of rotation in terms of a relative velocity, leaving the relativity of

---

[23] For a concise account of the bucket thought experiment (as well as other thought experiments of Newton and Leibniz), see (Arthur 2017).

motion intact. Two opposite points on the rim of a wheel of a moving wagon, for example, would have a certain relative velocity to one another if the wheel were rotating, and this would be the same whether the velocities were computed with respect to the ground or to the rest frame of the wagon, or to any other frame of reference in uniform motion relative to the wagon. No reference to absolute space is needed.[24]

Ernst Mach was equally committed to the position that the motion of any given body “can only be estimated by reference to other bodies” (Mach 1919, 230): “If we take our stand on the basis of facts,” he wrote, “we shall find we have knowledge only of relative spaces and motions” (232).

> The universe is not *twice* given, with an earth at rest and an earth in motion; but only *once*, with its relative motions, alone determinable. It is, accordingly, not permitted us to say how things would be if the earth did not rotate. …
>
> Newton’s experiment with the rotating vessel of water simply informs us, that the relative rotation of rotating the water with respect to the sides of the vessel produces *no* noticeable centrifugal forces, but that such forces *are* produced by its relative rotation with respect to the mass of the earth and the other celestial bodies. No one is competent to say how the experiment would turn out if the sides of the vessel increased in thickness and mass till they were ultimately several leagues thick. (Mach 1919, 232).

Einstein interpreted this as a positive conjecture of Mach’s, and one necessitated by his commitment to the relativity of motion and the inefficacy of a space independent of bodies. “In a consistent theory of relativity,” he wrote in 1917, “there can be no inertia relative to ‘space’, but only an inertia of masses *relative to each another*” (Einstein 1917, 180); quoted from Torretti (1996, 197). The basic idea is suggested by the Equivalence Principle. The centrifugal force driving the water up the sides of the bucket is the result of its inertial tendency to keep going at a tangent to the circumference of the bucket, and (in the rest-frame of the bucket) the slanted side allows it to creep upwards as a result of the consequent acceleration away from the centre. Extending this to the Earth, we can detect its rotation from the small centrifugal force that results from its spinning on its axis, even though it appears to be at rest while the stars are rotating through the night sky. But if, as Mach had argued, all motion is relative—even circular motion—then the same effects should occur if the Earth is really stationary and the rest of the mass in the universe is moving around it. Einstein took this to be licensed by the EP. For the same acceleration outwards could be produced by a suitable gravitational field, symmetrically disposed around the Earth and bucket, and this field would be the result of the rest of the mass in the universe.

Einstein was encouraged in this interpretation by his successful application of the EP to predict both the warping of spacetime and gravitational time dilation using the example of a rotating disc, which we discussed earlier. That thought experiment had taught him that the accelerative effects of rotation would be equivalent to those produced by an appropriate gravitational field. So he was led to model Mach’s response to Newton by supposing that the inertial effects in the bucket (including the centrifugal acceleration) could be caused by a thick spherical shell (modelling the

---

[24] For a discussion of Huygens’s views on motion, see Stein’s (1977), esp. pp. 8–9.

mass in the rest of the universe) surrounding the Earth and bucket, and producing an equivalent gravitational acceleration within the shell.[25] He assumed a Minkowskian spacetime, and calculated how the field due to the shell would distort the spacetime inside it. But (after several failed attempts) he found that the curvature produced was much too small to account for the curvature of the water in the bucket. Worse, he had calculated this by treating the effect of the shell as a small perturbation on the metric field of Minkowski spacetime, thereby implicitly assuming the boundary conditions for that spacetime. Moreover, most of the metric field is due not to the shell, but to the otherwise empty Minkowskian spacetime, thus violating Mach's principle.

The basic error here is the one described in the previous section: Einstein has mistaken the geometrical redescription of a physical situation for a physically inequivalent situation where acceleration is transformed into a gravitational field with its own source. That is, you can regard the acceleration of a body as physically produced by a gravitational field only if there is a source producing precisely that field. The covariance of the equations of GR does indeed allow the redescription of the bucket spinning relative to the Earth and shell to one using a frame of reference where the metric is rotating and the bucket at rest. And this can be reinterpreted as being due to a gravitational field. But this field does not have the shell as its source. Moreover, the values of the components of the field tend to infinity as we go outwards from the bucket. This is not physically equivalent to a field whose source is the non-rotating shell and whose values always remain finite, as Einstein had assumed in attributing all the inertia of the water in the bucket as due to the gravitational acceleration produced by the massive shell.

Thus there is no universal relativity of all motion, including rotational motion. In one respect, this should already have been obvious from a consideration of Copernicanism, as urged by Leibniz. If one body (say, the Earth) is spinning relative to another (the universe, represented by Einstein's shell), then even granting an equivalence between descriptions of the phenomena taking the Earth as spinning or at rest, this does not make for a physical equivalence: the equivalence of hypotheses is broken by the consideration of causes. There is clearly a huge asymmetry in what it would take to cause the Earth to spin and what it would take to make the universe spin around it, in terms not only of the energy required to produce each, but also in resultant effects.

This point was obscured for Einstein by his identification of the relativity of all motion with the covariance of the equations of general relativity, which guaranteed the equivalence of descriptions of the same phenomena in any coordinate system, however accelerated. This interpretation was bolstered by his conviction that all motion had to be motion with respect to other ponderable bodies, and his identification of coordinate systems as representing such bodies. From this Machian perspective, special relativity is defective in that the geodesics of Minkowski spacetime are defined independently of bodies. It is true that you could interpret Newton's bucket as rotating not relative to empty space, but relative to a body (or to any of an

---

[25] In a way, Einstein's reasoning is the obverse of Huygens's: where for Huygens gravity was a difference in centrifugal force, Einstein is making apparent centrifugal force an effect of gravity.

equivalence class of bodies) moving inertially, that is, along timelike geodesics in Minkowski spacetime. But these geodesics are given independently of the bodies in them. In general relativity, on the other hand, the geodesics are determined by the distribution of matter, so it is not out of the question that rotating matter would twist them so as to produce the effect of a centrifugal force in a stationary bucket. As we have seen, however, although it turns out that there is a small effect of this kind, this is mainly produced by the rotation of the metric field in an otherwise empty space, thus violating the Machian intent.[26]

In actual fact, GR provides the means to define rotation in a local neighbourhood in a way that does not depend on the frame of reference or on the distribution of matter. Consider a particle tracing its own worldline, and define a triad of mutually perpendicular spacelike vectors, like the thumb and first two fingers of a hand. Together with the 4-velocity vector of the particle these define a tetrad constituting a "compass of inertia" along the geodesic, with respect to which the local rotation of any small volume of fluid can be determined (cf. Torretti 1996, 200–201).[27] Physically this would manifest itself in the vorticity of the fluid. Now (in the spirit of Einstein) we may suppose such a rotating volume to be the body of an "observer", with a notion of "same time across the universe" defined by that observer's coordinate system. With a little more technical propriety, we could try to define such a "rotating observer's time" as a hypersurface orthogonal to the worldline of a rotating particle. But here we run up against an insuperable obstacle. As Dennis Dieks has explained in detail, "it turns out to be a basic characteristic of the rotating frame that the locally defined Lorentz frames do not mesh: they cannot be combined into one frame with a globally defined standard simultaneity" (Dieks 2004, 11). This is essentially the same point made by Torretti earlier with his example of Mannheim and Manhattan: we can impose a local Lorentz chart (spacetime coordinatization) onto a rotating particle, but we cannot in general extend it to the whole of spacetime. This is already the case in SR, and it goes over into GR spacetimes.[28] We can represent whether a system is rotating or not; but if we take a hypersurface orthogonal to the worldline of a rotating particle, "it will quite generally not lead to a global definition of simultaneity" (Dieks 2006, 162). There is no globally defined present associated with an arbitrary "observer". This consideration fortifies the conclusion we reached in discussing special relativity: there is no world-wide now that we can associate with an arbitrary observer in the universe.

Might it not still be possible, however, to conceive a global now even while ceding that time flows along timelike world lines that are parametrized by proper time? This is the conception advanced by the noted physicist George Ellis. If becoming is tracked by the lapse of proper time of all individual processes, he reasons, then the

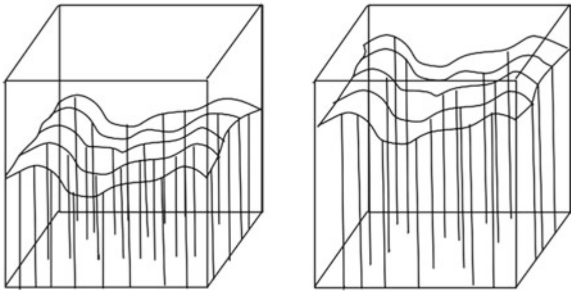
---

[26] There is also an analogous effect known as "frame-dragging", a warping of spacetime outside a massive shell produced by its rotation.

[27] It is in this sense—around the local compass of inertia—that the universe itself can be said to rotate, as in Gödel's rotating universe solution to the EFE.

[28] "These Lorentz frames at various events on the hypersurface do not mesh to form a global inertial frame, but their surfaces of simultaneity do mesh to form the spacelike hypersurface itself" (Misner et al. 1973, 714).

Fig. 7.3 Ellis's EBU



universe as a whole—the sum total of events that have occurred as of now—must be increasing in size. Thus he conceives the *now* as constituted by the current endpoints of the worldlines of all the processes that are occurring: "The ever-changing surface S(τ) separating the future and past—the 'present'—at the time τ is the surface {τ = constant} determined by the integral (20) along a family of fundamental world lines starting at the beginning of space time" (9). In Ellis's Fig. 2, this surface is pictured as a kind of uneven blanket held up by the poles of world lines extending back to the Big Bang (Ellis 2012, 4; see Fig. 7.3). "Spacetime grows," he writes, "as time inexorably evolves: at each new instant every previous present has become part of the past" (2012, 2–3). This results in a conception of becoming in terms of the uneven advance of a global now into an empty future, a view Ellis calls the Emerging Block Universe (EBU).

Attractive as such a view may seem at first, however, the EBU cannot be sustained. For it is a version of the expanding block view that I criticized in Chap. 3. To picture spacetime or the worldlines of processes as "growing" is to treat them as a quasi-spatial objects that can change. But that would require a time outside spacetime in which such change could be represented, whereas the 4-dimensional picture already includes time. Just as nothing moves along a worldline, so the worldline itself does not grow by the accretion of events; rather it is a four-dimensional representation of a process as if it has already occurred. There is no standpoint from which we can say that spacetime as a whole is either is changing or changeless, even though there is change everywhere in it. The classical conception of becoming in terms of the advance of a world-wide now must be discarded.[29]

But if that is the case, in what sense can the universe have a well-defined age? How can it be said to be some 13 to 14 billion years old? This is one of the main issues to be treated in the following section, concerning the global structure of spacetime.

---

[29] I can think of two further objections to Ellis's "uneven blanket" conception of the present as a surface of equal proper time elapsed. One stems from the Twin Paradox discussed in Chaps. 5 and 6: for Astrid, less of her proper time will have elapsed since leaving Terence than will have for him, yet at their reunion they are certainly present to one another. Second, the worldlines of bodies that have penetrated the Schwarzschild radius of a black hole will not terminate on any such surface of constant proper time.

The other is the issue of closed timelike worldlines and the possibility of time travel in a curved spacetime.

7.5 Relativistic Cosmology and Cosmic Time

One may well wonder what sense it makes to talk about the age of the universe in a relativistic setting. For that would seem to require that the notion of the universe at an instant is well defined. But as we have seen, according to the special theory there is no one inertial reference frame privileged over the others: we could as well take the time coordinate of a frame co-moving with a comet receding from us at near the velocity of light as take a time in a frame in which the Earth is at rest. In any case, the Earth is rotating, and we have just seen that for a rotating body no such local Lorentz chart can be extended across the whole of spacetime. Again, even if we disregard such an extension of local time across the whole cosmos and concentrate on proper time, the length of time some process has been taking will depend on its path though spacetime, so no two processes taking different paths to reach us from some distant point will have been in existence for exactly the same duration. So how can we talk of the universe having an age, as cosmologists do when they describe the universe as being some 14 billion years old?

To understand this, we need to take into account another remarkable intervention by Einstein. For his innovative application of General Relativity to the universe as a whole in 1918 precipitated developments that had the unintended (and unanticipated) consequence of allowing an age to be applied to the universe as a whole, thus reintroducing a notion of cosmic time by the back door, as it were. To set the scene for this remarkable turn-around, we need to say something about the state of cosmology at that time.

At the beginning of the twentieth century the question of the age of the universe was a matter of pure speculation, and beyond the pale of science. It's not that no progress had been made since the seventeenth century, when Newton had estimated its age at somewhat over 6000 years, based on "revealed knowledge" and Biblical chronology—essentially around thirty years for each generation since Adam. In fact, some intimations of geological "deep time" were already consequent on his contemporary Nicolas Steno's correct explanation of fossils: as was gradually appreciated, not only would there have to have been vast geophysical changes to the Earth's crust to explain how it is that sea creatures' fossils are found on what are now mountains, it would also have taken far longer than a few thousand years for mineral deposits to gradually replace the organic matter in the creatures that had become fossilized.[30] Also, it was commonly assumed by other scientists of Newton's day (such as Huygens and Leibniz) that the Earth had begun life as a fireball, and had then cooled to its present state. This allowed for estimates of a much longer life, such

---

[30] Leibniz had effectively recognized this in his posthumously published *Protogea*. See (Arthur 2014) for a discussion.

as the 75,000 years that Buffon calculated in 1779, on the basis of seeing how long it would take a globe of similar composition to cool. At the end of the nineteenth century, William Thomson (later Lord Kelvin) used his unsurpassed knowledge of thermodynamics to calculate ages for the Earth and Sun of about 20 million years, although even 100 million years appeared too short to geologists like Charles Lyell. In fact, Thomson's estimates were assumed by many to refute Darwin's theory of evolution, until the discovery of the radioactivity of Earth's core gradually led to a revision of its age to today's estimate of some 4 and a half billion years. But that was just the Earth; and how the Earth fitted in with the rest of the known universe was largely unknown. As late as the 1920s it was still a matter of dispute whether the universe just consisted of our own galaxy, with a few exotic spirals at the poles, or whether these spirals were galaxies like our own at a much greater distance from us.

All this changed beginning with Einstein's application of GR to cosmology in 1918, even if its significance for observational cosmology was not fully appreciated until the late 1920s and 30s. For it transpired that it followed ineluctably from certain solutions of Einstein's field equations—despite his initial disinclination to accept this—that space itself would not remain constant in volume, but must necessarily either expand or contract. Given this, together with Hubble's empirical confirmation that it is indeed expanding, it is possible to define a time coordinate that tracks the expansion back to its origin, some 14 bya. And it was Einstein himself who unwittingly precipitated this development by daring to apply his theory to the whole universe, although when he did so he assumed that it would be static and unchanging in its overall spatial dimensions.[31]

Einstein's route to this development was not motivated by problems of cosmology, but rather was precipitated by his dissatisfaction with Schwartzschild's assumption that the Sun's gravitational metric would converge to a flat Minkowskian metric at infinity. In one sense this was a sensible enough idealization: a free particle travelling in an otherwise empty space at a very great distance from the Sun should be travelling along a geodesic in a flat spacetime. But if this held at infinity, it would be in conflict with Mach's Principle, as understood by Einstein, for there would be no masses in an empty space at infinity to determine its geodesic. As Einstein wrote,

> In a consistent theory of relativity there can be no inertia *relative to* "*space*", but only an inertia of masses *relative to each other*. Hence if I take a mass sufficiently far away from other masses in the world its inertia must fall down to zero. (Einstein 1917, 145)

But the Dutch astrophysicist Willem de Sitter had in 1916 found solutions to the EFE for a spacetime devoid of matter, thus showing that general covariance was insufficient to guarantee compliance with Mach's Principle. Einstein—after some false steps taken to get around this with the aid of Jakob Grommer—had a brainwave: the problem of boundary conditions at infinity could be done away with if the universe were finite! Space could have a finite volume, but be curved in such a way that it had

---

[31] In my discussion here I am particularly indebted to Roberto Torretti's wonderfully learned book (1999) and article (2000), as well as to the excellent chapters by Michel Janssen (2014) and Christopher Smeenk (2014) in the *Cambridge Companion to Einstein*.

no boundary. Here the 2-D analogue would again be the surface of a sphere: you can travel as far as you like in any direction along the surface without any obstruction, yet the surface has a finite area. By analogy, a 3-D sphere could be curved in such a way that you could travel as far as you like in any direction through it without any obstruction, yet the sphere would still have a finite volume (remember, the curvature of such a space can be defined intrinsically—we do not need to regard it as contained in a 4-D spatial volume.) Accordingly, Einstein sought a solution of the EFE in which matter, represented on a large enough scale as a homogeneous and isotropic fluid, would yield a spacetime consisting in a sequence of finite spacelike slices of this 4-D volume having constant curvature. He could not find a solution, but found that he could if he added a new term to the left hand side, a constant λ times the spacetime metric $g_{\mu\nu}$:

$$R_{\mu\nu} - {}^{1}\!/_{2} R g_{\mu\nu} + \lambda g_{\mu\nu} = (8\pi G/c^4) T_{\mu\nu} \tag{7.10}$$

Here Einstein was guided by the requirement that his theory should reduce to the Newtonian theory of gravity in a suitable limit, and argued that the same gambit would also solve a difficulty for that theory, one that he believed had remained unsolved. The difficulty is that if all matter attracts all other matter, then, despite the weakness of the gravitational force at large distances, it looks as though all the matter in the universe, given enough time, should eventually collapse inwards. Adding the constant (which could be regarded as analogous to a constant of integration that had previously been arbitrarily set to zero) would have the effect of producing a subtle but universal force of repulsion to offset gravity, and thus preserve an equilibrium and constant volume.

This λ is Einstein's famous *cosmological constant*. Although he later repudiated it, declaring it his "greatest blunder", it has undergone a renaissance of late given the discovery that the rate of expansion of the universe is increasing.[32] But its revival in the 1990s is another story, which we need not pursue here. The significance of the 1918 paper is that it marks the beginning of relativistic cosmology by showing the way to apply GR to the universe as a whole. The essential assumptions Einstein had made in order to derive his solution were that spacetime admits a global time coordinate x_0, and that perpendicular to the geodesics parametrized by x_0 are spacelike hypersurfaces, having Riemannian metrics of constant curvature—essentially, that spacetime could be treated as having a 3 +1 dimensional metric, 3 of space and one of time. In Friedmann's (1922) and (1924) the Russian mathematician Alexander Friedmann used these assumptions to derive a family of solutions to the modified EFE of (6.10) that are spatially homogeneous and isotropic—that is, such that matter is uniformly distributed in it and that there is no preferred direction in space. On this model, the universe is represented as a pressureless fluid whose worldlines are along the geodesics parametrised by x_0. The distinctive feature of Friedmann's solutions is that the worldlines do not remain at a constant mean distance from one another as

---

[32] It is worth remembering, even so, just how very small the constant is. Current estimates put it at $1.11 \times 10^{-52}\ \mathrm{m}^{-2}$!

in Einstein's static universe, but either diverge from one another or converge (or else diverge then converge).

Friedmann's solutions were largely ignored at the time. This was partly because Einstein claimed they were based on a mathematical error (a criticism he soon retracted), but also because they seemed to contradict the "obvious" fact of the universe's constancy. It was not until they were rediscovered by the Belgian priest Georges Lemaître in his (1927) and championed by Eddington (1930) that they really gained traction. For Lemaître drew attention to the fact that the observations of Hubble and Shapley were indicating that the universe was not in fact static, and to Hubble's publication in his (1929) of the linear relationship between the velocities of spiral nebulae and their distances it became apparent that this was predicted by the EFE on the basis of Friedmann's and Lemaître's solutions.[33] Einstein's own solution yielding a static universe (like that of Willem de Sitter, its main rival at the time) was shown by Eddington (1930) to be a special case that was also unstable: like a tennis ball perched on top of a basketball, the slightest nudge would see it tumble one side or the other, into expansion or contraction.

So now this could be added to the three principal pieces of empirical evidence in favour of GR. In addition to the anomalous precession of Mercury, the deflection of light as it passes the Sun, and the gravitational red shift, all predicted by Einstein and supported by Schwartzschild's solution, here was the impressive and unexpected implication of the non-static condition of the universe, empirically confirmed as an expansion. It is the latter that has crucial relevance for time and becoming. For in legitimizing the notion of cosmic time in which the universe expands, it has been thought to allow for the reintroduction of the classical conception of becoming in terms of hyperplanes of simultaneity. So James Jeans argued in his (1935), and it was against his notion of cosmic time that Gödel launched his CTC-containing solutions as a counter-example. It is also of the utmost relevance to the arguments for the elimination of time in some canonical quantum theories of gravity, since it is this cosmic time that is eliminated in such theories.

How is this cosmic time established? As we have seen, it depends on making some drastic idealizing assumptions about the universe as a whole in order to obtain a set of linear equations that can then be solved for it. Friedmann construed it as a powder devoid of stars, galaxies, clusters or any large scale structures.[34] In a very clear review of relativistic cosmology in 1933 H. P. ("Bob") Robertson refers to this as "this rawest of all possible approximations" (Robertson 1933, 64). The approximation rests on two empirically supported assumptions: (1) that even though extra-galactic nebulae appear grouped in clusters, the clusters themselves appear uniformly distributed throughout the cosmos, like molecules in an ideal gas, and

---

[33] In fact Hermann Weyl had predicted the linear relationship later empirically confirmed by Hubble in his (1923), and it was H. P. Robertson's endorsement of Lemaître's prediction of the expansion of the universe on this basis in his (1928) that was confirmed by Hubble in his (1929) and to Hubble's (1923).

[34] Robertson (1933, 64). Although Friedmann had assumed it was pressureless, this assumption was dropped by Lemaître, Robertson and Walker.

(2) that their relative motions are small compared with their mean motion.[35] The first of these assumptions is now called the “Cosmological Hypothesis”. The second depends on an assumption first explicitly identified by Hermann Weyl in 1923, that “there exists in each region of cosmic space-time a *mean* motion which represents the actual motions to within relatively small and unsystematic deviations”, now known as “Weyl’s Postulate” (Robertson 1933, 64). It entails, as Robertson explains, that “the world lines of all matter belong to a pencil of geodesics which converges toward the past”. Each of these “pencils” represents the worldline of an extra-galactic nebula at rest in a comoving reference frame at cosmic time t equal to the proper time of any co-moving system. In his (1935) Robertson showed using group theory and symmetry considerations that any solution satisfying the Cosmological Hypothesis and the Weyl Postulate would be of the kind derived by Friedmann and Lemaître. Similar results were obtained by Walker (1935), so that all such models are now known as Friedmann-Lemaître-Robertson-Walker, or FLRW spacetimes. Given the linear relationship between distance and velocity reported by Hubble, this allowed an extrapolation backwards to a time when the mutual distance of the nebulae is almost zero, the era of the so-called Big Bang.

Now, this might be thought to be a slender basis for such a big claim as that the universe originated only a finite time ago. Why, for example, could it not have been in an unstable equilibrium for an unspecified time before tumbling into expansion? As a matter of historical fact, the simple model just outlined was challenged on many fronts, not least of which was its initial prediction of a universe only 2 billion years old, but paradoxically containing stars considerably older. This was revised as distance measurements were improved (especially with the recalibration of the distances of so-called Cepheid variable stars), but for some decades it vied with several rival theories, including the Steady State Theory of Hermann Bondi, Thomas Gold, Fred Hoyle and others, according to which the expansion is constantly fuelled by the creation of hydrogen atoms everywhere throughout spacetime, with no origin of time needed or supposed. The FLRW spacetime model did not become firmly established as the standard model of the cosmos until after the discovery of the cosmic background microwave radiation by Penzias and Wilson in their (1965). For this served as an unanticipated confirmation of a model for the generation of the light elements (heavier than hydrogen but no heavier than iron) at very great densities (such as would have existed in the universe some 13 bya), which would have resulted in a massive explosion of radiation permeating the universe, evenly in all directions. Subsequent expansion would have cooled this radiation to about 2.7 K, and this is the microwave radiation that Penzias and Wilson had discovered to be uniformly incoming from all directions with the signature of blackbody radiation produced in

---

[35] Gödel also remarked on the lack of precision in the notion of a “mean motion of matter”: “What may be called the ‘true mean motion’ is obtained by taking regions so large that a further increase in their size does not any longer change essentially the value obtained. In our world this is the case for regions including many galactic systems” (Gödel 1949, 559, n. 7). He adds that this approximation could perhaps be improved, but would still involve “introducing more or less arbitrary elements (such as, e.g., the size of the regions or the weight function to be used in the computation of the mean motion of matter)” (560, n. 9).

a hot early universe predicted by the Big Bang model. With the subsequent fusion of particle physics and cosmological theory, further refinements have been made, including the hypothesis of a period of cosmic inflation, according to which the expansion of the universe went into a massive overdrive, increasing in size by a factor of at least 10^{26} in roughly the first 10^{-36} to 10^{-33} s after the Big Bang.

To what extent, though, can all this be regarded as a way of recovering "the intuitive idea of an absolute time lapsing objectively", as James Jeans proposed (Jeans 1935, 22–23)? The first point to be made is that this is hardly a recovery of the classical notion of simultaneity. There is no sense in which the hyperplanes at different values of the cosmic time function correspond to any kind of physical connectivity among distant events. The lack of inherent meaning attaching to the simultaneity of distant events that Mermin saw as "the single most important lesson to be learned from relativity" (Mermin 2005, xii) is even more pertinent here, where we are talking about galactic systems that are so far apart that their mutual motions are negligible compared to the mean common motion due to expansion. The cosmic time tracking that expansion has no bearing at all on local becoming.

Moreover, the lapse of time is tracked by the proper time along the geodesics of any of these co-moving nebulae, not by the coordinate time associated with the local motion of any physical system therein. But this itself raises some very pertinent questions about the appropriateness of extrapolating backwards to the beginning of time, and about time and becoming in the earliest epochs. For the existence of a cosmic time function, and with it the applicability of the FLRW spacetime metric, holds only so long as Weyl's Postulate can be maintained. Imagine, then, the geodesics extrapolated back to when they were all much closer together. There would have been much more interaction of matter under the effect of gravity because of this proximity—a veritable maelstrom of condensing and exploding stars, with concomitant forming of black holes and quasars, etc. And so time lines would be more twisted locally than now. But the situation is much more problematic as we extrapolate back to times closer to the supposed initial singularity, as has been emphasized by Rugh and Zinkernagel (2009), (2017). Setting aside the quantum effects that would be associated with the Planck time at 10^{-43} s, and before the hypothesized era of inflation at around 10^{-34} s, severe doubts about the validity of Weyl's Postulate can be entertained for much earlier periods in the backwards extrapolation. Even if there were no galaxies in the very early universe, there would at least have been protons, neutrons and pions in bound states that might feasibly have had a mean motion. But prior to the quark–hadron phase transition that is postulated to have taken place at around 10^{-5} s, there would have been no physical systems in bound states; and prior to about 10^{-11} s, the so-called Higgs phase transition, there would have been no particles with mass (i.e. non-zero rest mass) at all, so no set of timelike worldlines would even be identifiable.[36] One may certainly doubt, therefore, whether it is reasonable to extrap-

---

[36] "Problems with the notion of a global cosmic time may arise if a privileged set of world lines becomes difficult to identify, e.g. in the very early universe above the electroweak (Higgs) phase transition or in a (complicated) inhomogeneous universe. A more serious problem for time (which is a problem even for a local definition of time) arises if a point is reached in the backward extrapolation

olate cosmic time back into such early epochs, or whether it makes sense to talk of time then at all.

The earliest stages of the universe are naturally of concern to those theorists concerned with quantum gravity (as we will discuss further in Chap. 8). Lee Smolin and Roberto Mangabeira, for example, have argued for the necessity of a global time in order for us to talk about the evolution of the universe, and also in order to support their suggestion that the laws themselves are not eternal, but emerge at a certain point in the history of the universe (and may yet change!). Smolin does not rest his case for a global time on cosmic time, however, preferring the formulation of general relativity called *shape dynamics*. This provides a preferred foliation of spacetime into three-dimensional spaces of constant mean curvature, evolving in a "dynamically determined preferred global time" (Unger and Smolin 2015, 420). Smolin defends the apparent incompatibility of such global nows with special relativity by arguing that, since these nows of global time "would be determined by the dynamics of the universe as a whole", they "would thus not be determinable in terms of information local to an observer" (420).[37] In other words, the existence of such a global time would be unobservable, and would make not a jot of difference to the local becoming of any processes, even on a galactic scale.[38] How far back such a 3 +1 dimensional model could extend towards the beginning of such a time is another question, which I will not attempt to tackle here.

Gödel posed a different problem for the definition of cosmic time by deriving solutions for Einstein's Field equations in which no cosmic time function is definable. The solutions he found were for a rotating "cylindrical" universe. Granted, our own universe does not appear to be rotating, but his point was that the very existence of such solutions shows that the existence of a cosmic time function is a contingent matter.[39] The remarkable feature of his solutions is that they allowed closed timelike curves (CTCs), worldlines traced by accelerating matter that curve right back to their starting point in spacetime. These are not geodesics, and it would take a massive amount of energy to accelerate along such a path, but Gödel's point is one of principle: if the spacetime point at which the path begins is the same as the one at which it ends, time cannot be said to have lapsed. If this is even possible in principle, he argued, then time is unreal.

---

where the world lines themselves can no longer be identified. In particular, this appears to be the case if some point is contemplated, e.g. at the onset of inflation, where all constituents of the universe are of a quantum nature, leading to what can be called the *quantum problem* of time." (Rugh and Zinkernagel 2017, 379).

[37] There is also the defence that the existence of a preferred foliation in a particular solution does not in itself constitute a violation of Lorentz invariance or covariance. As Ellis points out, "These are symmetries of the general theory, not of its solutions. Interesting solutions break the symmetries of the theory" (Ellis 2012, 10). The same point is made by Kent Peacock (2018, 120).

[38] Smolin posits that "all theories of subsystems of the universe should be relativistically invariant so that the effects of the preferred global time, if it exists, are not detectible in experiments at less than the scale of the whole universe" (Unger and Smolin 2015, 491–2).

[39] Here Gödel seems to be conveniently forgetting that in GR the very structure of spacetime is contingent, since it depends on the distribution of matter, as has been argued by Dorato (1995, 283). So the existence or not of CTCs is necessarily contingent.

Of course, Gödel is assuming that becoming must consist in "an infinity of layers of 'now'" which come into existence successively" (Gödel 1949, 557), as we saw in Chap. 5, which is why he targets the cosmic time championed by Jeans. But this cosmic time has no bearing on the becoming of local processes. Gödel's argument for CTCs, in fact, presupposes such local becoming of the very processes tracing them.[40] Their existence would still be paradoxical, however, since a process tracing one of them should have done so in a finite lapse of proper time, integrated along its path; yet the proper time between coincident points on a worldline should be 0.

It is not my intention to give a thorough analysis of the philosophical implications of CTCs here.[41] But the preceding discussion brings into play an important consideration about the identity of events in relativistic spacetimes which we need to address. It is usually assumed in this context that two events are identical if they occur at the same spacetime point.[42] This is because in relativity theory, spacetimes are not inert containers within which events can be located. The structure of any spacetime region is dynamically determined by the distribution of mass and energy in it. Schwartzschild's solution is a good example of this, showing how spacetime is curved around a single massive, spherically symmetric body. But in the FLRW and Gödel solutions the universe is modelled in an highly idealized fashion, with all the inhomogeneities and irregularities characteristic of realistic processes simply ignored. Only gravity is included in the model, and even here, under highly idealized assumptions—for instance that on a sufficiently large scale, the worldlines of clusters of galaxies can be modelled as though they were worldlines of dust or of a perfect fluid. Once we suppose more realistic physical processes tracing Gödelian timeloops, however, we are faced with acute problems of self-identity of events and processes.

Suppose, for example, we imagine a bat hitting a ball at a certain spacetime point. Perhaps the bat cracks under the impact, so that its worldline from that event forward is that of a cracked bat. Now suppose that the event occurs in a spacetime allowing closed timelike curves, and the bat is carried around such a curve right up to the moment of impact. If the curve is completed, what is the state of the bat immediately on impact with the ball? It cannot be both cracked and not cracked. Similar problems will occur with any process that bears the marks of its history—which means practically any realistic process at all.

Such considerations suggest that the idealized conditions assumed in the FLRW and Gödel solutions will break down, although how they do so in such a way as to preclude the formation of time loops is not known, and will depend on the detailed physics of the cases. Various so-called "causality conditions" have been proposed that would prevent the occurrence of CTCs, among them the *Strong Causality Condition*,

---

[40] I argued this in Arthur (1982); see also Dorato (1995) and (Dorato 2006), Savitt (2009) and Dieks (2006).

[41] For such a thorough philosophical discussion see Arntzenius and Maudlin (2013).

[42] For example, Robert Geroch begins his analysis by defining the identity of an event as follows: "We regard two events as being 'the same' if they coincide, that is, if they occur 'at the same place at the same time.'" (Geroch 1978, 4).

which stipulates that any point in spacetime exists in a neighbourhood through which at most one timelike curve can pass. This condition, also called the *Chronology Principle*, is tantamount to requiring an Alexandrov topology at every point in spacetime, thus grounding the requirement of chronological precedence, which I have taken as a precondition for the reality of (local) process. So we may see this logic as reversing Gödel's argument. Where he argued that the existence of CTCs refutes the reality of (global) becoming (albeit while presupposing the reality of local processes), one can instead take the requirement of the reality of local processes with continuous local histories to preclude the existence of CTCs.

7.6 Summary

- In this chapter we explored some of the main implications of Einstein's theory of General Relativity for the understanding of time. In GR the shape of spacetime is effected by the distribution of matter and its motions, and issues concerning passage are now inseparable from the treatment of gravitation. Three important manifestations of this are the phenomenon of gravitational time dilation; the extreme warping of spacetime near massive black holes and its implications for temporal passage; and lastly the possibility of closed time-like curves (CTCs), and the threat they pose to the reality of becoming.
- In Sect. 7.2 Einstein's Strong Equivalence Principle was introduced, and used to explain the curvature of spacetime, the phenomena of time dilation and the temporal behaviour of objects travelling into black holes. Although inertial motion is conceived differently in GR than in SR, proper time still has the same role of tracking time lapsed along the worldlines of processes.
- In Sect.7.3 we proceeded to an account of the origins of GR, in order to clarify confusions about what general relativity actually consists in. Einstein viewed it as the universal relativity of all motions, variously construing it as embodied in an equal validity of all coordinate systems for describing the phenomena guaranteed by the Equivalence Principle, in the requirement of the covariance of the equations of the theory, and in Mach's principle. We saw that these attempts foundered on mistaking the geometrical redescription of a physical situation manifested in GR for a physically inequivalent situation where acceleration is transformed into a gravitational field with its own source. So GR does not endorse a complete physical equivalence of reference frames in the general sense that Einstein desired to capture, where time is simply relative to the observer's coordinate system. Moreover, the principle of local becoming is embodied in the geodesic principle of GR, which guarantees the same connection of time with inertia as was ensconced in Newton's physics and preserved in SR.
- Section 7.4 consisted in a consideration of Mach's Principle and the relevance of rotation for the understanding of space, time and spacetime. It was seen that local rotations cannot be transformed away by choosing an appropriate reference frame without unwanted physical consequences, so that GR does not (contrary to

Einstein's intentions) support a universal relativity of motion. It was argued that the very existence of these local rotations makes it impossible in GR (or in SR) to define a global time coordinate by extending the time coordinate of a rotating system across the whole of spacetime. Thus GR makes the arguments for the locality of becoming only more compelling.

- Finally, in Sect. 7.5 we examined issues concerning the status of becoming in relativistic cosmology. We saw that Einstein's application of GR to the universe as a whole resulted in the establishing of the FLRW solutions, which explained the (spatial) expansion of the universe and allowed the definition of a cosmic time. But contrary to the claim of Jeans that this allows the revival of a global becoming, it was argued that the expansion of the universe occurs on a cosmic scale, with no obvious bearing on local passage. Gödel's solutions to the field equations of GR for a rotating cylindrical universe introduced the prospect of CTCs, which he used to argue against the reality of becoming. It was argued that Gödel's argument presupposes that becoming must be global, and that the requirement of the reality of local becoming may be construed as a sufficient condition for precluding a spacetime containing CTCs.

References

Almheiri, A., D. Marolf, J. Polchinski, and J. Sully. 2013. Black Holes: Complementarity or Firewalls?. Preprint at http://arxiv.org/abs/1207.3123(2012).

Arntzenius, Frank, and Tim Maudlin, 2013. Time Travel and Modern Physics. In *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), ed. Edward N. Zalta. https://plato.stanford.edu/archives/win2013/entries/time-travel-phys/.

Arthur, Richard T.W. 1982. Exacting a Philosophy of Becoming from Modern Physics. *Pacific Philosophical Quarterly* 63 (2): 101–110.

Arthur, Richard T.W. 2013. Leibniz's *Mechanical Principles* (c. 1676): Commentary and Translation. *Leibniz Review*, 23: 101–116.

Arthur, Richard T.W. 2014. *Leibniz*. Cambridge: Polity Press.

Arthur, Richard T.W. 2017. Thought Experiments in Newton and Leibniz, 111–127. In *The Routledge Companion to Thought Experiments*, ed. James Robert Brown, Yiftach Fehige, and Michael T. Stuart. Palgrave Macmillan.

Barbour, Julian. 2011. Shape Dynamics. http://arxiv.org/abs/1105.0183v1 [gr-qc] 1 May 2011.

Dieks, Dennis G.B.J. 2004. Space, Time and Coordinates in a Rotating World. In *Relativity in Rotating Frames*, ed. G. Rizzi, and M.L. Ruggiero. Kluwer: Accessed through Open Access.

Dieks, Dennis. 2006. Becoming, Relativity and Locality. In *The Ontology of Spacetime*, ed. Dennis Dieks, 157–176. Amsterdam: Elsevier.

Dorato, Mauro. 1995. *Time and Reality. Spacetime physics and the objectivity of temporal becoming*. Bologna: CLUEB.

Dorato, Mauro. 2006. The Irrelevance of the Presentist/Eternalist Debate for the Ontology of Minkowski Spacetime. In Dieks, ed. 2006, 93–109.

Dyson, F.W., Arthur S. Eddington, and C. Davidson. 1920. A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society*. https://doi.org/10.1098/rsta.1920.0009.

Eddington, Arthur. 1930. On the Instability of Einstein's Spherical World. *Royal Astronomical Society. Monthly Notices* 90: 668–678.

Einstein, Albert. 1905. Zur Elektrodynamik bewegter Körper. In *Annalen der Physik*, vol. 17, 1, 891–921; CPAE 2, Doc. 23; English translation in Lorentz et al. 1923, 37–65.

Einstein, Albert. 1907. Über das Relativitätsprinzip und die gezogenen Folgerungen. *Jahrbuch für Radioactivität und Elektronik* 4: 411–62. CPAE 2, Doc. 47.

Einstein, Albert. 1911. Über den Einfluß der Schwerkraft und die Ausbreitung des Lichtes. *Annalen der Physik* 35, 898–908. CPAE 3, Doc. 23; English translation in Lorentz et al. 1923, 99–108.

Einstein, Albert. 1912. Relativität und Gravitation. Erwiderung auf eine Bemerkung von M. Abraham. *Annalen der Physik* 38, 1059–64; CPAE 4, Doc. 8.

Einstein, Albert. 1917. Kosmologische Betrachten zur allgemeinen Relativitätstheorie, CPAE 6, Doc. 43; translated as "Cosmological Considerations on the General Theory of Relativity" in Lorentz et al. 1923, 177–188.

Einstein, Albert. 1918. Dialog über Einwände gegen die Relativitätstheorie, *Die Naturwissen–schaften* 6: 697–702, 29 November; an English translation exists on the web at http://en.wikisource.org/wiki/Dialog_about_objections_against_the_theory_of_relativity.

Einstein, Albert. 1919. Time, Space and Gravitation. *The Times*, November 28, 1919; CPAE 7, Doc. 26; reprinted as "What is the Theory of Relativity?" in (Einstein 1954), 227–232.

Einstein, Albert. 1921. On the Theory of Relativity, 247–249 in (Einstein 1954).

Einstein, Albert. [1922] 1956. *The Meaning of Relativity*. Fifth Edition. Princeton: University Press.

Einstein, Albert. (CPAE). 1987. *The Collected Papers of Albert Einstein,* ed. J. Stachel et al. Princeton: Princeton University Press.

Einstein, Albert and Marcel Grossman. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner. CPAE 4, Doc. 13.

Einstein, Albert, and Leopold Infeld. 1938. *The Evolution of Physics*. New York: Simon and Schuster.

Einstein, Albert. 1954. *Ideas and Opinions*. New York: Crown Publishers.

Ellis, George F.R. 2012. Spacetime and the passage of time. http://arxiv.org/abs/1208.2611v4 [gr-qc] 26 Aug 2012.

Friedmann, Alexander. 1922. Über die Krümmung des Raumes. *Zeitschrift für Physik* 10: 377–386.

Friedmann, Alexander. 1924. Über die Möglichkeit einer Welt mit konstanter Åegotiser Krümmung des Raumes. *Zeitschrift für Physik* 21: 326–832.

Geroch, Robert. 1978. *General Relativity from A to B*. Chicago: University of Chicago Press.

Gödel, Kurt. 1949. A Remark about the Relationship between Relativity Theory and Idealistic Philosophy. In *Albert Einstein: Philosopher-Scientist*, ed. Paul Arthur Schilpp 557–562. New York: Tudor.

Hubble, E.P. 1929. A relation between distance and radial velocity among extragalactic nebulae. *Proceedings of the National Academy of Sciences* 15: 168–173.

Huygens, Christiaan. 1690. *Discours de la Cause de la Pesanteur*. Leiden: Pierre: Vander AA, Marchand Libraire.

Isaacson, Walter. 2007. *Einstein: His Life and Universe*. New York: Simon and Schuster.

Jaki, Stanley L. 1978. Johann Georg von Soldner and the Gravitational Bending of Light, with an English Translation of His Essay on It Published in 1801. *Foundations of Physics*, 8 (11/12): 927–950.

Janssen, Michel. 2014. 'No Success like Failure…' Einstein's Quest for General Relativity,1907–1920. 167–227 in (Janssen and Lehner 2014).

Janssen, Michel and Christoph Lehner, eds. 2014. *The Cambridge Companion to Einstein*. New York. Cambridge University Press.

Jeans, James. 1935. *Man and the Universe*, Sir Halley Stewart Lectures.

Leibniz, Gottfried Wilhelm. 2001. *The Labyrinth of the Continuum: Writings on the Continuum Problem, 1672–1686*. Ed., sel. & transl. with an introductory essay. R. T. W. Arthur. New Haven: Yale University Press.

Lemaître, G. 1927. Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale de nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles* 47A: 49–59.

Mach, Ernst. 1919. *The Science of Mechanics*. Transl. Thomas J. McCormack. 4th edition. Chicago/London: Open Court.

Mermin, N. David. 2005. *It's About Time: Understanding Einstein's Relativity*. Princeton and Oxford: Princeton University Press.

Misner, Charles W., Kip S. Thorne and John Archibald Wheeler. 1973. *Gravitation*. San Francisco: W. H. Freeman and Co.

Muller, Richard A. 2016. *Now: The Physics of Time*. New York: W. W. Norton.

Newton, Sir Isaac. 1730. *Opticks: or, a Treatise of the Reflections, Refractions, Inflections, and Colours of Light* (4th ed., corrected. London: William Innys.

Norton, John D. 2018. The Hole Argument. In *The Stanford Encyclopedia of Philosophy,* ed. Edward N. Zalta (Summer 2018 Edition). https://plato.stanford.edu/archives/sum2018/entries/spacetime-holearg/.

Pais, Abraham. 2005. *Subtle is the Lord: the science and the life of Albert Einstein*. Oxford: Oxford University Press. 1982; reissued with a new foreword by Sir Roger Penrose.

Peacock, Kent A. 2018. *Quantum Heresies*. Milton Keynes, UK: College Publications.

Penrose, Roger. 2005. *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Vintage Books.

Penrose, Roger. 2016. *Fashion, Faith and Fantasy in the New Physics of the Universe*. Princeton and Oxford: Princeton University Press.

Penzias, A.A., and R.W. Wilson. 1965. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophysical Journal* 142: 419–421.

Pound, R.V., and G.A. Rebka. 1960. Apparent Weight of Photons. *Physical Review Letters* 4: 337–341.

Pound, R.V., and J.L. Snider. 1965. Effect of Gravity on Gamma Radiation. *Physical Review* 140B: 788–803.

Robertson, H.P. 1928. Relativistic Cosmology. *Philosophical Magazine* 5: 835–848.

Robertson, H.P. 1933. Relativistic Cosmology. *Reviews of Modern Physics* 5: 62–90.

Robertson, H.P. 1935. Kinematics and World Structure. *Astrophysical Journal* 82: 284–301.

Rugh, Svend E., and Henrik Zinkernagel. 2009. On the Physical Basis of Cosmic Time. *Studies in History and Philosophy of Modern Physics* 40: 1–19.

Rugh, Svend E., and Henrik Zinkernagel. 2017. Limits of Time in Cosmology. In ed. Khalil Chamcham, Joseph Silk, John D. Barrow and Simon Saunders *The Philosophy of Cosmology*, 377–395. Cambridge University Press.

Ryckman, Thomas A. 2014. Early Philosophical Interpretations of General Relativity. In ed. Edward N. Zalta, *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). http://plato.stanford.edu/archives/spr2014/entries/genrel-early/.

Savitt, Steven F. 2009. The Transient *nows.* In *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony,* ed. W. Myrvold and J. Christian, 339–352. (Springer-Verlag).

Smeenk, Christopher. 2014. Einstein's Role in the Creation of Relativistic Cosmology, 228–269 in (Janssen and Lehner 2014).

Stein, Howard. 1977. Some Philosophical Prehistory of General Relativity. In *Foundations of Space-Time Theories,* ed. Earman, John, Clark Glymour and John Stachel, 3–49 Wisconsin: Minnesota University.

Torretti, Roberto. 1996. *Relativity and Geometry*. New York: Dover Publications. (Corrected reprint of Torretti, *Relativity and Geometry*, Oxford: Pergamon Press, 1983).

Torretti, Roberto. 1999. *The Philosophy of Physics*. Cambridge: Cambridge University Press.

Torretti, Roberto. 2000. Spacetime Models for the World. *Studies in History and Philosophy of Modern Physics* 31 (2): 171–186.

Unger, Roberto Mangabeira, and Lee Smolin. 2015. *The Singular Universe and the Reality of Time*. Cambridge: Cambridge University Press.

Voltaire. 1733. *Letters Concerning the English Nation*. London: C. Davis and A. Lyon.

Walker, A.G. 1935. On Riemannian Spaces with Spherical Symmetry about a Line and the Conditions for Isotropy in General Relativity. *Quarterly Journal of Mathematics* 6: 81–93.

Weyl, Hermann. 1918. "Gravitation und Elektrizität", *Königlich Preußische Akademie der Wissenschaften* (Berlin). Sitzungsberichte. 465–80.

Chapter 8
Becoming in Quantum Theory

> *The quantum universe is static. Nothing happens; there is being but no becoming. The flow of time and motion are illusions.*
> —Julian Barbour (2008, 2).

8.1 Introduction

In the preceding chapters of this book we have been considering the status of becoming in the context of classical mechanics and spacetime theories. Different issues arise in a quantum theoretical context. The chief difference is this: whereas relativity theory begins by assuming a manifold of events on which its spacetime structures are defined, in quantum theory events are more nearly the *results* of quantum processes. *That* they so result, and with the probabilities predicted by the theory, is not in question. But exactly *how* events are supposed to arise is one of the main difficulties of interpretation of the theory. Let me attempt to set the scene with an anecdote that will introduce one of the main themes.

The celebrated Argentine writer (and self-styled "amateur metaphysician") Jorge Luis Borges wrote an ingenious short story, "The Garden of Forking Paths". In it he introduces his readers to the universe of Ts'ui Pên, who "believed in an infinite series of times, in a growing dizzying net of divergent, convergent and parallel times" (Borges 1964, 53). This network of times, Borges writes, "which approached one another, forked, broke off, or were unaware of each other for centuries, embraces *all* possibilities for time." As the protagonist in the story discovers, this conception of a garden of forking time paths is not just the key to understanding the baffling novel Ts'ui Pên bequeathed to his heirs, it is the universe in which he himself is living.

It is also the universe in which *we* live, if we are to believe some modern quantum physicists. Here the reference is to the Many Worlds Interpretation of quantum theory (the MWI or "multiverse") pioneered by Everett III (1957) and further developed by DeWitt (1970), with Neill Graham (DeWitt and Graham 1973), and others. This inter-

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_8

pretation has been embraced with enthusiasm particularly by quantum cosmologists and the pioneers of quantum information theory.

Some flavour of the appeal of branching time and the multiverse can be gleaned from its utilization by David Deutsch and Michael Lockwood to resolve the "Grandfather Paradox" of time travel provoked by the possibility of closed time-like curves in spacetime (CTCs) discussed in the previous chapter (Deutsch and Lockwood 1994). Their time-traveller, Sonia, traversing such a CTC, returns to the year 1934 and persuades her grandfather not to marry her grandmother. This is a paradox, as their not marrying (nor, presumably, conceiving one of her parents out of wedlock!) prevents her from ever having come into existence. Deutsch and Lockwood propose a resolution of this and similar paradoxes by enlisting the multiverse interpretation of quantum theory. For "one thing we already know about CTCs," they say, "is that we need quantum mechanics to understand them" (1994, 72). Invoking Everett's interpretation, according to which "if something physically can happen, it does—in some universe" (1994, 72), they conceive many universes, linked together in one convoluted spacetime. The links among these universes "force Sonia to travel to a universe that is identical, up to the instant of her arrival, with the one she left, but that is thereafter different because of her presence" (1994, 73). She may succeed in preventing her grandfather from marrying her grandmother, but this will occur in a different universe from the one she left, thus eliminating any contradiction in her life history.

This resolution of the grandfather paradox makes redundant the supposition of the Chronology Principle defended in the previous chapter. Sonia's worldline in the hypothetical scenario described by Deutsch and Lockwood does not violate the principle of the continuity of identity along any given worldline. On completing the CTC, Sonia has retained her self-identity in a self-consistent way along her own unique worldline—only, the universe she returns to is not identical with the one she left! The two universes are identical in all details prior to the time in 1934 when she visits her grandparents, but after that they diverge. Now, such a multiplication of universes might seem a steep price to pay for not having to embrace the Chronology Principle—but this is the least of the extravagance required by the MWI, which requires such a splitting into multiple distinct copies at each instant at which there is a multiplicity of possibilities according to quantum theory, and for any continuous spectrum of possibilities (like the time of a radioactive decay), this requires an uncountable infinity of copies![1]

What, then, prompts this interpretation, what problems does the MWI resolve in quantum theory itself, and why is it particularly popular among quantum cosmologists and quantum information theorists? To answer such questions we first need to introduce the basic tenets of quantum theory, and discuss the various difficulties in understanding what it says about the world. This I attempt in Sect. 8.2, where I give

---

[1] Although I will not pursue this further, Deutsch's hypothesis seems objectionable for other reasons. For this scenario presumes a time common to the two universes that Sonia exists in. But if she is present in one and not the other, and the structure of spacetime depends on the masses in it, we cannot presume a common time for the two universes.

a brief general description of how quantum theory originated in difficulties besetting the classical theory in accounting for the interaction of matter with radiation. I then describe the matrix mechanics of Werner Heisenberg and the rival theory of wave mechanics of Erwin Schrödinger, and how they were reconciled in Max Born's probability interpretation. Schrödinger had proposed an equation giving the evolution of the state ψ of a system in time, and had interpreted ψ as a wave amplitude. But Born showed that ψ should instead be interpreted in terms of probability amplitudes whose squares gave the probabilities for the transitions described by Heisenberg's matrix theory. The wave- or state-function evolves continuously in time according to Schrödinger's equation, but it can be expressed in terms of a superposition of "eigenstates", corresponding to the occurrences of various outcomes or "eigenvalues", each outcome occurring with a corresponding probability. On measurement, however, only one outcome occurs, with a probability corresponding to one of the probability amplitudes. This was thought to occur by the discontinuous "collapse" of the wave function into one eigenstate on measurement. That idea was given a formal rendering by John von Neumann, who proposed that quantum theory involved two different kinds of process: the continuous evolution of the wave function, and a discontinuous process of collapse on measurement, involving the projection of the wave function representing the various possible outcomes onto the eigenstate representing the actual outcome (his Projection Postulate). The ensuing controversy concerning the status of this postulate and the difficulties attendant on it are known collectively as the "measurement problem".

In fact, quantum theory was from the beginning embroiled with problems of interpretation. The projection postulate was devised by von Neumann in 1932 in support of the Copenhagen Interpretation developed by Niels Bohr, where it was interpreted as a formalization of Bohr's contention that physical quantities have no independent reality prior to being brought into being by observation or measurement. Heisenberg had articulated that view earlier, supporting it with his famous "Uncertainty Principle" (to be explained below). But the Copenhagen Interpretation was opposed by Einstein and Schrödinger, to whom the idea that observation precipitates events appeared as a preposterous irrealism, and they each devised thought experiments to articulate their objections. To this end Schrödinger constructed his famous Cat Paradox thought experiment, and also diagnosed the entanglement of quantum states as *the* characteristic feature of quantum theory.[2] This feature, according to which systems that have interacted continue to have their component states interlinked after they are spatially separated, was also exploited by Einstein in his famous EPR thought experiment contained in a paper co-written with Boris Podolsky and Nathan Rosen.

It is not part of my brief here to try to resolve all the issues involved in interpreting quantum theory. Friends and colleagues whose views I respect take many opposing stands, and a proper treatment of such matters is in any case beyond the scope of a book of this kind. My aims are (at least, ostensibly!) more modest. I aim only to treat those issues bearing on time in quantum theory, and what I have to say about

---

[2] Schrödinger (1935a, 555). For an illuminating discussion, see Jammer (1974, 211–218).

the various interpretations of the theory will be restricted to a consideration of their implications for the reality of becoming.

One of these, introducing the topic of Sect. 8.3, is the contention of Bondi, Reichenbach and Ellis that quantum theory, through its indeterminism, allows a rigorous notion of becoming to be recovered in terms of transitions from possibility to actuality. A potential objection to this is that such discontinuous transitions depend on accepting the projection postulate. The status of that postulate, as noted, has been hotly disputed. Many regard it as an unnecessary addendum to quantum theory, lacking proper foundation in its theoretical apparatus, and have therefore rejected it. Everett's MWI is a prime example of such a "no-collapse" interpretation. According to this view, there is no collapse of the various quantum possibilities into actualities on interaction with a measuring apparatus (or any other physical system): rather they are all actual. It is just that we experience only one of them, being, necessarily, on only one branch, in a branching multiverse. I argue that this has the effect of inverting the measurement problem: instead of accounting for becoming in terms of a transition to the actual, it presupposes a computationally intractable matrix of actualities, and bequeaths us problems in accounting for the different probabilities of occurrence predicted by quantum theory. A proper understanding of the objective chance that is entailed by quantum theory, I contend, not only makes the positing of the projection postulate unnecessary, but also undermines the motivation for the actualist interpretation of possibilities inherent in the MWI and casts doubt on the possibility of a wave function for the entire universe.

A second line of response to the measurement problem, however, takes us to the question of the treatment of time in a relativistic quantum context, the subject of Sect. 8.4. Here we consider an interpretation that is more directly heir to the criticisms of Einstein and Schrödinger, arising out of the search for "hidden variables" that would reinstate a realist ontology for quantum particles underpinning the statistics of quantum theory. This was the approach of Louis de Broglie, revived by David Bohm, and championed and elaborated by John Bell for the best part of three decades. First proposed by Bohm as providing an easily visualizable resolution of the EPR paradox, it is embroiled with the whole issue of quantum non-locality. The quantum statistics of measurement results are sensitive to changes in experimental set-up, and on the de Broglie-Bohm view, this sensitivity entails a non-local, causal, reconfiguring of possible particle trajectories. Moreover, Bell proved a theorem which, he argued, proved that any theory that attempted to account for correlations between previously interacting systems that are now remote from one another would violate the (empirically confirmed) quantum statistics if it did not allow possible superluminal influences between them. I argue that the correlations involved cannot be interpreted as processes occurring across space-like distances, not only because no processes can occur over spacelike intervals (as argued in Chap. 6), but also because the correlations are among probability amplitudes, not events.

In the last section, Sect. 8.5, I turn to time in the context of a quantum theory of spacetime. According to fundamental quantum principles, the theory of general relativity must break down in very small regions of spacetime. This has precipitated the search for a theory of quantum gravity that would preserve the basic principles of each theory, or at least show how they could be derived from a putatively more general theoretical framework. I shall not be exploring here all such approaches, but only those approaches which are held to necessitate the elimination of time. In particular, I examine the arguments of two of the main protagonists of this view, Julian Barbour and Carlo Rovelli, and subject them to criticism.

8.2 The Basics of Quantum Theory

Quantum theory had its origins in attempts to resolve certain anomalies in classical thermo-dynamics at the beginning of the twentieth century, anomalies concerning the interaction of radiation and matter that seemed to defy the predictions of the classical theories concerned. One of these concerned the radiation emitted by a "black body", that is, a cavity with a black internal surface that is in thermal equilibrium with the matter and radiation it contains. Two formulas had been derived for the relationship between the intensity of the radiation emitted and its frequency at a given temperature: the *Rayleigh-Jeans Law* and *Wien's Law*. The first worked well at low frequencies, but diverged to infinity at the high end of the spectrum, while the second worked well at high frequencies but not at low ones. Max Planck discovered that he could "fix" the anomaly by a purely ad hoc stratagem, adding as an otherwise unmotivated constraint on the classical equations the condition that the energy E be exchanged between matter and radiation only in discrete bundles or "quanta" that are integral multiples of the frequency ν,

$$E = h\nu,\ E = 2h\nu,\ E = 3h\nu, \text{etc.}$$

Here h is Planck's constant, having the dimensions of *action*, the physical quantity first proposed by Leibniz. Its dimensions are $[\mathrm{ML}^2\mathrm{T}^{-1}]$, equivalent to the product of energy and time, or of length and momentum, or of angular momentum and angle. Planck's "fix" thus amounts to requiring that action can only come in integral multiples of h, hence the term "quantization of action". In 1905 Einstein took Planck's idea for a fundamental fact rather than an adventitious fix, and used it to solve a second anomaly concerning the interaction of matter and radiation, the *photoelectric effect*. When high frequency radiation was used to bombard a metal so as to dislodge electrons, it was expected classically that increasing the intensity of the radiation would lead to an increase in the energy of the electrons. Instead, it was found experimentally that increasing the intensity only led to increasing the number of electrons produced, while their energy depended only on the energy of the incident radiation. Einstein showed how this could be accounted for if the incident radiation is emitted or absorbed by the electrons only in discrete quanta of energy $h\nu$, as if it consisted in particles, which were later dubbed "photons".

The success of this realistic interpretation of Planck's quanta then became the foundation of what is called the Old Quantum theory, a set of ad hoc proposals centred on the quantization of action. Following Einstein's lead, Niels Bohr solved another classical anomaly. According to Maxwell's theory, negatively charged electrons in orbit around a positive atomic nucleus should continuously emit radiation and fall into the centre, thus collapsing the atom. To explain why they did not, Bohr proposed a model of the atom in which the orbits of the electrons were constrained to be consistent with the quantization of action, so that radiation would be emitted in discrete energy quanta as the electron fell from one orbit to another, but could not fall below the bottom orbit corresponding to the minimum quantum of action.

The particulate nature of light in its interaction with matter was extended by Einstein. Combining Planck's Law, $E = h\nu$, with his famous $E = mc^2$ from Special Relativity, he proposed that radiation would possess not only mass but momentum, $h\nu/c$ or h/λ (since $c = \nu\lambda$). This was confirmed in 1923 by Arthur Compton's experiments with gamma rays and electrons, where the rays collided with the electrons as though they were particles with momentum h/λ. de Broglie (1927) then extended Einstein's reasoning. If light could behave like a particle when it interacted with matter, perhaps matter could also behave like light? Then anything at all possessing a mass m moving at a velocity v would have a corresponding wavelength $\lambda = h/mv$. Because of the tiny value of Planck's constant h, 6.63×10^{-34} J s^{-1}, this would not be noticeable for macroscopic objects, but for elementary particles of correspondingly tiny mass such as the electron of mass (9.11×10^{-31} kg), wave behaviour should be observable. This was confirmed by the experiments of Davisson and Germer in 1923, where beams of high-speed electrons passing through crystals of nickel showed diffraction patterns typical of waves with wavelength $\lambda = h/mv$.

On Bohr's model an electron could be excited to a higher energy level in the atom by absorbing a light quantum in a discrete event, but both emission and absorption of light quanta would defy classical description in terms of continuously orbiting particles and continuously distributed waves in space. Noting this, in 1925 Werner Heisenberg developed an interpretation that eschewed classical trajectories for particles altogether, and gave a mathematical treatment of energy transitions solely in terms of observable quantities. This was his *matrix mechanics.* At the same time, Erwin Schrödinger was developing De Broglie's conjecture into a quite different approach to the same enigma, his *wave mechanics*, which he published in a series of papers in 1926. Taking off from the Bohr model where the electrons could be interpreted as electric current distributed in standing waves, Schrödinger thought of the electron orbits as "beat phenomena": like the standing waves you can produce by sending waves down a skipping rope, only orbits corresponding to whole numbers of wavelengths would be allowed. Finding difficulty with producing a Lorentz-invariant treatment, he began with a Galilean-relativistic model whose *pièce de résistance* was the famous Schrödinger equation:

$$\hat{\mathrm{H}}\psi(t) = i\hbar\frac{\partial}{\partial t}\psi(t)$$

Here $\hbar$ is the so-called reduced Planck constant, $h/2\pi$, and $\hat{H}$ is an operator based on the classical Hamiltonian function representing the total energy of the system in terms of the field potentials to which it is subject. $\psi(t)$, the ψ-function, represents the wave as a function of the time t. To apply the equation, terms representing the kinetic and potential energy of the system are inserted into the Hamiltonian, and the resulting partial differential equation is solved for the wave function. For certain potential functions, such as those describing the Bohr atom, the equation predicts the quantization of various physical quantities, in this case the angular momentum characterizing the different Bohr orbits.

Schrödinger derived the equation on the model of the classical Poisson wave equation, and interpreted it as describing real waves propagating in space. He followed de Broglie in interpreting quanta like Einstein's photons as "wave packets" propagating in space, bunched-up waves that would travel with a group velocity (somewhat analogously to how a traffic snarl on the highway due to a traffic accident travels away from the scene of the accident towards approaching vehicles). But it was soon realized that highly localized wave packets would disperse extremely quickly, in a mere fraction of a second, and therefore could not represent particle trajectories.[3] The narrower the wave packet, and therefore the better localized it is, the greater the spread in values of the wavelengths of the waves from which it is composed, and the greater its tendency to disperse. This is an instance of Heisenberg's "Uncertainty Principle", according to which there is a trade-off between the spread in wavelengths (each corresponding to a momentum of $p = h/\lambda$) of the component waves, and the width of the packet, so that the spreads in momentum p and position x are related by $\Delta p \cdot \Delta x \cong h$. (We will give a more precise statement and discussion of this principle below.) A second decisive objection to Schrödinger's wave interpretation is that there is not one ψ-function for each quanton[4] constituting a system, but one such function representing them all. If each quanton were a wave in ordinary three-dimensional space, then a system of n quantons would be n waves combining in this space. Instead, however, the ψ-function in Schrödinger's equation applies to all n systems together in an abstract $3n$-dimensional space called *configuration space*. This would not be a problem if the constituent waves retained their independence as they do classically. But in quantum theory even if one begins with n independent subsystems with their own distinct ψ-functions, on interacting they become entangled and are no longer separable. So a wave in configuration space of $3n$ dimensions no longer corresponds to n waves in ordinary 3-space, as it does in the case of a single quanton (when $n = 1$).

The conundrum of how to interpret Schrödinger's wave function was resolved in 1926 by Max Born (who, along with Pascual Jordan, had helped Heisenberg develop matrix mechanics). He proposed that $\psi(t)$ is a *probability amplitude*: taking its square gives the probability density, and by means of it one can determine the probability of a given physical quantity being found to have a particular value (its "expectation value"). So Schrödinger's equation describes not the propagation of waves in real

---

[3] In terms of the traffic analogy, a snarl steadily moving away from the accident towards approaching vehicles disperses as cars approaching the location take exits before they get there. But a wave packet of atomic dimensions (about 10^{-10} m) disperses in space so quickly that it would grow to about a km wide after 1 μs!

[4] The term "quanton" is due to Mario Bunge, and denotes an entity like an electron, photon or atom, that exhibits typical quantum behaviour.

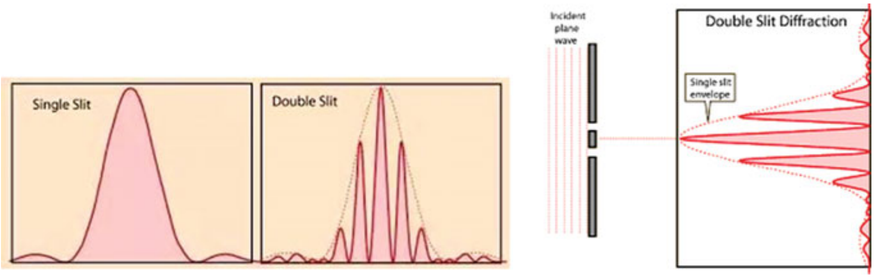


Fig. 8.1 The two-slit experiment

space, but of information concerning the physical quantities characterizing the quantum system, with the propagation occurring in configuration space. But that is rather too abstract an explanation. We can get a grasp of what it means physically with the help of a concrete example, that of the famous two-slit experiment.

Here we imagine a stream of quantons fired at a screen. It does not matter whether they are electrons, as in an old television, where the picture is produced by an electron gun firing electrons to produce light flashes on the screen, or photons of light, as we will take them to be here. A light wave encountering an obstacle or passing through a slit is able to bend round the corners, so that light appears in the geometrical shadow of the obstacle. This effect is called diffraction. It can be modelled by imagining numerous point sources of light that overlap at certain points beyond the obstacle or slit, and cancel each other out at others. This produces a diffraction pattern, a big light patch in the middle, and light and dark fringes in the geometrical shadow. The effect is enhanced when there is more than one slit of the same size suitably close. (This was the effect exploited by Davisson and Germer in their experiments demonstrating the wave-like behaviour of electrons with a crystal lattice.) With two slits thus arranged, we get the characteristic interference pattern pictured above in Fig. 8.1, where the bright middle region of the diffraction pattern is broken up into alternating light and dark patches.[5]

On the classical picture of a wave propagating in space, this is understandable, as the wave is in a medium filling the space, and the motions of its particles can be seen to combine or cancel one another as the wave fronts propagate (as was originally shown by Huygens in the seventeenth century). As we have seen, though, the quantum nature of light means that its interaction with matter is by one quanton (photon) at a time. If a photon is thought of as a particle with a well-defined trajectory, how could it pass through both slits at once and interfere with itself? Now, one might persist in visualizing light as streams of such particles which could exhibit wave behaviour by their group motions, like the molecules in a water wave. The intensity of a light wave, however—as Einstein had shown—does not depend on the energy, but only on the number of photons. So it is possible (at least in the case of a laser beam) to reduce

---

[5] Illustration from http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/mulslid.html#c2, accessed January 29, 2019.

the intensity of light to such a degree that only one photon is detected at a time. And yet, as the pattern builds over time, one photon at a time, the same diffraction and interference effects are produced! Moreover, if one of the slits is closed and the experiment is repeated, no interference pattern results.

These latter facts are particularly puzzling if light is conceived as a stream of photons propagated in distinct trajectories through space and time. With one or the other slit open, flecks of light are produced on the screen, building up to the large bright central patch of the diffraction pattern as more photons reach the screen. But when the second slit is opened, flecks of light appear in places where there would be none if either one of the slits were open alone, and the bright patches are up to four times brighter than any bright patches produced when one slit is closed.[6]

But of course one can delay the decision to close one of the slits until after the photon should have passed the slits. Such a delayed choice on the part of the experimenter then seems to determine whether the light has passed through only one slit, or both, combining to form an interference pattern. And again, the intensity of the light can be reduced so that photons are detected at the screen one by one. So in such *delayed choice experiments*, the experimenter's choice seem to determine which path the photon has taken through space: through both slits or through only one. Such considerations were instrumental in persuading Bohr to develop the Copenhagen Interpretation, according to which there are no phenomena prior to their being brought into being by observation or measurement.

Indeed, John Archibald Wheeler (who had worked with both Bohr and Einstein) took such delayed choice experiments as providing a clinching argument for Bohr's interpretation and against Einstein's reservations. In support he devised a delayed choice thought experiment in a cosmological setting (Wheeler 1983), appealing to the phenomenon of gravitational lensing, where the mass of an intervening galaxy bends the light coming from a distant quasar (in accordance with Einstein's GR). This results in two images of the same quasar reaching the observer as the light comes around opposite sides of the galaxy. Since the light is detected one photon at a time, the detector can be set up to detect from which direction the photon came. Alternatively, by inserting a half-silvered mirror (his "½S") at the point where the rays cross, they can be combined to form an interference pattern, producing a bright spot at detector II if the two rays are in phase with one another, and a dark one in detector I, where no photon is ever detected. (see Fig. 8.2.) So, depending on the experimenter's choice, it can be determined which route the photon took, or whether, so to speak, it came "by both routes" to interfere at the detector. However, Wheeler observes, "at the time the choice was made whether to put in ½S or leave it out, the photon in question had *already* been on its way for billions of years." (Wheeler 1983, 193). It is therefore "wrong to talk of the 'route' of the photon" at all, Wheeler concludes, repeating the Bohrian mantra that "No elementary phenomenon is a phenomenon until it is a registered (observed) phenomenon".[7] (193).

---

[6] Penrose gives a typically lucid account of the 2-slit experiment in his (2016, 20–23).

[7] Wheeler (1983, 184, 193); he cites Bohr's (1958, 73, 88).

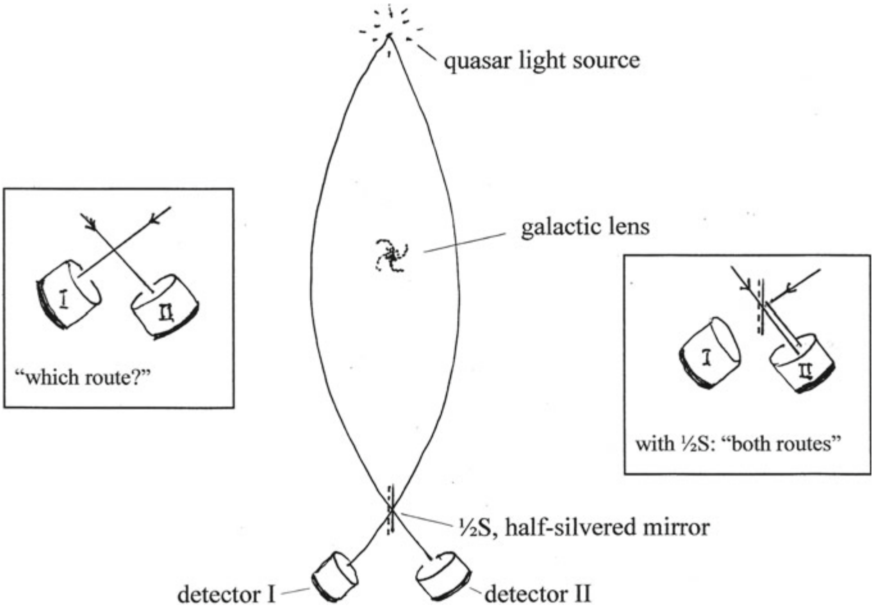


Fig. 8.2 Wheeler's delayed choice experiment

But Wheeler interprets the implications of the delayed choice experiment in a way that goes well beyond Bohr, and connects directly with the issue of the reality of becoming. "What we have the right to say of past spacetime, and past events," he argues, "is decided by choices—of what measurements to carry out—made in the near past and now" (194). In this way, he thinks, events in the past are actually *created* by the act of observation or detection, the past having "no existence until it is recorded in the present" (194). So it is not so much that it is wrong to ascribe a route to the photon, as that the route does not become a phenomenon "until it has been brought to a close by an irreversible act of amplification" (193). That is, he interprets the experimenter's choice as determining the route that the photon actually took in the past. If the mirror ½S is left out, the photon from the quasar must be said to have gone around the galaxy on one side or the other; if it is inserted, then the photon has, so to speak, been made to come around on both sides. But prior to detection, there was no fact of the matter, despite the billions of years that must have elapsed since its supposed emission.

Wheeler's notion of the creation of the past is hardly compelling, though. It is only if we insist on treating the photon as a particle with a well-defined trajectory that we can infer its route retroactively. It is not even clear that his interpretation is in keeping with the Copenhagen Interpretation. Bohr would have held that the phenomenon in question is not the particle's route through spacetime, but the quantum of light's being recorded as absorbed at one detector or another. For him, as for Heisenberg, there is no fact of the matter concerning the photon's trajectory.[8] Moreover, we cannot tell from recording a *single* photon's absorption at detector II, say, whether it has come from one side only or from both, "interfering with itself" to arrive there. Still, it remains the case that when sufficiently many detections are made the statistics of the experimental results will differ depending on whether or not the experimenter inserts the mirror ½S: with the mirror in place, none are recorded at detector I, whereas roughly equal numbers will be recorded at each detector if the mirror is not inserted. This is puzzling, since the statistics are built up one event at a time, so that somehow the individual detection events must be sensitive to the experimenter's intervention.

Wheeler is by no means alone in imagining quantons as particles with well-defined trajectories (albeit with the usual qualifications about their wave behaviour). It seemed self-evident to Born,[9] and it underlies the de Broglie-Bohm theory we will be discussing further in Sect. 8.3, according to which the wave equation describes a *pilot wave* guiding particles selectively into regions where the probability density is highest. Analysing the two-slit experiment, J. S. Bell (one of the foremost protagonists of the theory) asks "Is it not clear from the smallness of the scintillation on the screen that we have to do with a particle? And is it not clear, from the diffraction and interference patterns, that the motion of the particle is directed by a wave?" (Bell 1987, 191). On this interpretation, the wave corresponding to a single particle is spread out through all of space, and encompasses all possible particle trajectories. When Wheeler's mirror ½S is interposed, the possible trajectories are instantaneously altered, and the particle follows one or the other of these from then on until it is detected only in the bright fringes at the screen.[10]

Bell was motivated, like Bohm, Einstein and Schrödinger before him, by a fervent opposition to the subjectivism implicit in the Copenhagen Interpretation that phenomena are brought into being by observation or measurement. That doctrine had received a strong initial impetus from Heisenberg's analysis in the 1927 paper in which he introduced his Uncertainty Principle. There he imagined using a gamma-ray microscope to observe an electron utilizing the Compton effect. The smaller the wavelength of the light employed to determine the electron's position, the greater the change in its momentum. "Thus, the more precisely the position is determined the less precisely the momentum in known, and conversely" (Heisenberg [1927] 1983, 64). He summarized this in the formula $\Delta x \cdot \Delta p = h$. Insisting that position can only be identified by definite experiments through which it can be measured,

---

[8] Cf Bohr (1958, 73): "[I advocate using] the word *phenomenon* to refer only to observations obtained under circumstances whose description includes an account of the whole experimental arrangement" (Wheeler and Zurek eds. 1983, 3).

[9] As Born writes in his autobiography: "I was witnessing the fertility of the particle concept every day in Franck's brilliant experiments on atomic and nuclear collisions and was convinced that particles could not simply be abolished" (Born 1968, 55; quoted from Torretti 1999, 336).

[10] If instead of a single particle there are N particles, then this change of the wave–function must take place in a 3N-dimensional abstract space, not physical space. But it will still have the effect that any of the particles are made to follow trajectories in real 3-dimensional space, since they must be detected there.

Heisenberg concluded from this that "the word 'path' has no definable meaning" (65), and that the electron's orbit "comes into being only when we observe it" (73). Bohr concurred, commenting that "the product of the inaccuracies in measurement of position and momentum are given by the general formula [$\Delta x \cdot \Delta p = \Delta t \cdot \Delta E = h$]" (Bohr [1928] 1983, 100). But as Schrödinger pointed out, this confuses two things, "the error distribution of the measurement, on the one hand, and the theoretically predicted statistics, on the other", two things that "have nothing to do with each other"(Schrödinger 1935b, 159).

The status of Heisenberg's principle was clarified by H. P. Robertson in 1929. He showed that what follows from the theory is that for a given variable A, its "uncertainty", ΔA, can be taken as the root mean square of its deviation from the mean (its expectation value $\hat{A}$). It follows that, if two variables A and B are Fourier transforms of one another—*canonically conjugate* variables, as they are called—such as position and momentum, these spreads about the mean are given by the formula $\Delta A \cdot \Delta B \geq \hbar/2$.[11] Thus Δx and Δp are not experimental errors, "uncertainties" in putative exact values of x and p due to the limitations of measurement, they are statistical scatters or spreads about the expectation value predicted by the theory, as evidenced by many runs of the same physical arrangement. This standard deviation interpretation of uncertainty, however, will not work for all physical situations, as been shown by Jan Hilgevoord, Jos Uffink and David Atkinson.[12] "A more adequate measure of the spread of a probability distribution," they suggest, "is the length W_α of the smallest interval on which a sizeable fraction α of the distribution is situated." (Hilgevoord and Atkinson 2011, 478). In sum, what Heisenberg's principle is telling us is that the more precisely a given physical arrangement determines, say, a quanton's position x within a certain range, the less determinate will be the range of values for its canonical conjugate variable p, and vice versa. It is therefore better called the Indeterminacy Principle, as recommended by Bohm (1957, 85 n.*).

Also, Heisenberg's and Bohr's endorsement of a similar uncertainty relation between time and energy has long been known to be in error: that a given quantum system should have a definite value of energy at a given time is a constitutive assumption of quantum physics, and clearly in violation of $\Delta E \cdot \Delta t \geq \hbar/2$ (even assuming the Δt could be given a definite meaning). Using Hilgevoord's formulation, though, an energy-time indeterminacy relation can be derived, for instance, for the energy spread of a quantum state of a quanton and its half-life $\tau_{1/2}$ (the time it takes for half of an ensemble of such quantons to decay): "if $\alpha = 0.9$, then W_α is

---

[11] Robertson proved that in general, if the commutator $[A, B]$ of the two Hermitian operators representing A and B is given by $i\hbar C/2$, then $\Delta A \cdot \Delta B \geq |C_0|\hbar/2$ where C_0 is the expectation value of C. For canonically conjugate variables, $C = C_0 = \pm 1$, yielding the above result (Robertson 1929, 127).

[12] "The standard deviation is not the most obvious, and certainly not the most adequate measure of uncertainty in quantum mechanics. For many perfectly normal quantum states the standard deviation diverges, and even when the wave-function approximates a δ-function, the standard deviation may remain arbitrarily large. A consequence of this fact is that [the Robertson inequality] permits probability distributions of p and q to be simultaneously arbitrarily narrow, contrary to what might be expected from an uncertainty relation." (Hilgevoord and Atkinson 2011, 487).

the smallest interval on which 90% of the energy distribution is situated", so that $\tau_{1/2}W_{0.9} \geq 0.9\,\hbar$ (Hilgevoord 1996, 1454).

Heisenberg's contention that physical quantities come into being only as a result of being observed led some physicists to propose that quantum theory essentially involves observers' consciousness. It was in opposition to such intrusions of subjectivism that Einstein and Schrödinger formulated their thought experiments, EPR and Schrödinger's famous cat. But to properly appreciate them we need first to introduce the measurement problem, which was a second significant factor in physics' subjectivist turn.

8.3 Quantum Indeterminism and the Measurement Problem

In Schrödinger's way of representing (Galilean-relativistic) quantum theory,[13] the state function of a given system in a certain environment does not give the precise values of position, energy, etc., that a system has at a given time. Rather it is a superposition of probability amplitudes, each corresponding to a given outcome for an identically prepared system: for instance, one such outcome might be the appearance of the system on the other side of a potential barrier. The probability of that event is calculated using the square of the corresponding probability amplitude. But if the system does actually manifest there (as detected, say, by a Geiger counter), we have an actual event. Quantum probabilities are replaced by actualities, indeterminacy by determinacy. It is this transition from the undetermined to the determined, according to one prominent interpretation, that constitutes becoming in physics.

This line of interpretation was advocated by Herman Bondi in a letter to the editors of *Nature* we already partially quoted in Chap. 4:

> … the flow of time has no significance in the logically fixed pattern demanded by deterministic theory, time being a mere coordinate. In a theory with indeterminacy, however, the passage of time transforms statistical expectations into real events. (Bondi 1952, 660)

Similarly, Hans Reichenbach, insisting that "if there is becoming, the physicist must know it", identified it with the transition from the undetermined to the determined:

> The concept of *becoming* acquires significance in physics: The present, which separates the future from the past, is the moment when that which was undetermined becomes determined, and 'becoming' has the same meaning as 'becoming determined.' … it is with respect to *now* that the past is determined and that the future is not. (Reichenbach 1956, 269, 270)

A similar line has been taken more recently by George Ellis: "the passing of time marks the change from indefinite (not yet existing) to definite (having come into

---

[13] This is called the *Schrödinger representation*: the operators are fixed, and the ψ-function varies in time. I will work in this representation, as opposed to the (equivalent) *Heisenberg representation*, where the ψ-function is fixed, and the operators vary in time.

being); the present marks the instant at which we can act and change reality." (Ellis 2014, 2).[14]

There are several distinct issues that need to be untangled in this connection. First there is the claim, articulated here by Bondi, that in a deterministic theory like Newtonian mechanics or relativity theory, there is no becoming, "time being a mere coordinate". We already treated this claim in Chap. 4, where we saw that becoming is entirely compatible with determinism, and is in fact presupposed by the very notion of process that is at its base. One persistent motivation for this mistaken claim has been the idea that free will requires indeterminism. But this, we saw, is based on faulty reasoning. On the contrary, a notion of free will worth defending requires the decision of the agent to be one of the chief determining factors in the cause. If quantum processes are *stochastic*—that is, statistically but not individually determined—then for that very reason they cannot constitute determinate requisites for individual conscious decisions.

Two other difficulties with the claims of Bondi, Reichenbach and Ellis should be clear from our discussions in Chaps. 5, 6 and 7: there is no world-wide now separating the future from the past; and Reichenbach's identification of time as a mere coordinate completely misses the special role of proper time in tracking the lapse of time in relativity theory.[15] This is symptomatic of a general tendency in the interpretation of quantum theory, as we shall see, to concentrate on Galilean-relativistic quantum theory, where the time coordinate is treated as if it had its classical significance, leading to results that are then found to be in conflict with Special Relativity.

What, then, is the contrast Bondi, Reichenbach and Ellis are alluding to? It would seem to be this. In classical physics, probabilities were employed when there was insufficient information about the events occurring to allow for a causal deterministic account to be given. For example, as we saw in Chap. 4, Laplace assumed as much in his approach to probability theory, as did Boltzmann in founding statistical mechanics. Even though the state of each particle at an instant is characterized by its velocity and position, it was not necessary to know the exact trajectory of each particle of an ideal gas in order to make determinations about probability distributions of the states of such a gas at any time. It was still assumed that in principle there was a fact about what the actual motions of the particles of the gas might be, so that at a micro-level everything would be consistent with causal determinism. The macro-level description in terms of probabilities reflected only a lack of knowledge of precise initial conditions.

But (on the orthodox account) the probabilities occurring in quantum theory cannot be interpreted in the same way. Although at first some quantum theorists subscribed to the "ignorance interpretation" of quantum theory—that its probabilities represented our ignorance of the precise causal trajectories at the micro-level—it was

---

[14] In his (2012, 9) Ellis defines the present as "the surface where the indeterminate future is changed to the definite past at any instant". As we saw in the previous chapter, Ellis conceives this surface as constituted by the endpoints of a family of world lines. I am indebted to Mauro Dorato for reminding me of Ellis's agreement with Bondi and Reichenbach on this matter.

[15] The latter objection does not apply to Ellis, who defines his global present in terms of proper time (but see the previous chapter for criticisms).

eventually realized that this was incompatible with orthodox quantum theory. The Heisenberg inequalities preclude the simultaneous specification at any instant of the position and momentum (and thus velocity) of a quantal system. (This means that the configuration space of n quantal systems is quite different from classical phase space, where each point in the space corresponds to a precise position and velocity for each subsystem.) Moreover, we are not dealing with classical probabilities, but with probability waves, which are propagated through time according to the quantum mechanical equations of motion. The wave amplitudes combine linearly, but it is their squares that are probability densities, resulting in cross-terms analogous to the '$2ab$' in $(a + b)^2$. The result is that the states are *entangled* with one another, and the probability waves corresponding to different constituent subsystems may produce interference (as happens in light waves), producing quite different statistical results than if they had consisted in a stream of classical particles, where the probability densities, not their amplitudes, would add linearly.[16] So probabilities are not reducible to ignorance of initial conditions: they represent objective indeterminacies, not epistemic ones.[17] The transition from a system's being in a certain region, say, with a given probability density, to its actually being detected in that region is what Bondi, Reichenbach and Ellis are regarding as the transition from the objectively undetermined to the determined.

Here, however, we confront a problem. On the accepted interpretation, the ψ-function of each "wave" or state is a linear superposition of terms, each one an eigenstate corresponding to an outcome appropriate to the experimental situation together with its probability amplitude, and this state function evolves continuously in time according to Schrödinger's equation. For instance, when a beam of quantons having two possible spin-orientations is deflected in an inhomogeneous magnetic field (as in the Stern-Gerlach experiments) the ψ-function would be expressed as a superposition of eigenstates $\varphi_{+½}$ and $\varphi_{-½}$, corresponding respectively to the eigenvalues spin-up (+1/2) and spin down (−1/2):

$$\psi = \alpha_{+1/2}\varphi_{+1/2} + \alpha_{-1/2}\varphi_{-1/2}$$

Here the probability of finding a quanton in the spin up state is $|\alpha_{+½}|^2$, and spin down $|\alpha_{-½}|^2$, each of these being a number between 0 and 1, inclusive.

---

[16] But it is also a mistake to regard the waves described by Schrödinger's equation as classical waves, as does J. S. Bell: "The mathematics itself is smooth, deterministic, 'classical' mathematics … of classical waves" (Bell 1987, 187; ellipsis his). Because of the phase factor, the wave amplitudes described by the psi-function do not transform in accordance with the Galilean transformations. The "puzzle" of wave-particle duality is not resolved by regarding a quanton as both wave and particle as Bell suggests (191), since it is not properly either a classical wave or a classical particle.

[17] Here, as we shall see in Sect. 8.3, exception must be made for the heterodoxy of Bohmian mechanics. In that theory there is a distinction between the dynamical variables whose values are subject to the Heisenberg inequalities, and the hypothesized variables determining the precise positions and trajectories on a subquantum level. The indeterminacies of the former result from imprecise knowledge of the initial conditions appropriate to the latter.

On measurement, however, only one outcome occurs, so that the system (assuming it has not been destroyed by the measurement) is now held to be in one of the eigenstates corresponding to a definite value of the observable in question—say, spin up with probability 1. The superposition has "collapsed", as it were, into just one of the eigenstates, with a probability corresponding to one of the probability amplitudes. Traditionally, this is known as "the collapse of the wave function", and the discrepancy between the continuous evolution in time of the state function and the discontinuous process of collapse, together with the difficulties of interpretation entailed by this discrepancy, is known as the "measurement problem".

This is something of a misnomer. It is not necessary that there be a measurement or an observing physicist for this to be a problem. All that is required for a "measurement" is that the quantum system interacts with its environment so as to produce some effect that is in principle observable. Bohr had insisted that this involves an "experimental apparatus" that is always classical—a pointer, for example, or a mark on a screen caused by amplification of a quantum event up to a level where it could be observed. But if the measuring apparatus is instead treated as a quantum system, then it would appear that after measurement the system should be in a state of superposition with the apparatus, not in an eigenstate. In a careful formal analysis, von Neumann (1932) showed that the same results would follow whether or not the apparatus is included in a quantum superposition, provided that there is a final stage in which there is a projection from the various possible outcomes onto the one actual outcome of the observation corresponding to the eigenvalue in question. He therefore proposed his Projection Postulate: in addition to processes of continuous evolution in time as described by the Schrödinger equation ("interventions of type 2"), there must be other discrete processes where the wave function of the combined system + apparatus collapses discontinuously into an eigenstate ("interventions of type 1"). Since any stage of the measurement prior to its being observed would involve material objects that could be included in the superposition without projection, von Neumann concluded that the projection itself must ultimately occur in the observer's consciousness.

Similar analyses were offered by London and Bauer (1983) in 1939, and by Eugene Wigner (1983) in 1962. Given the discrepancy between the mere potential of the quantum mechanical wave function and the definiteness of phenomenological experience, to them it seemed that the only way to avoid the arbitrariness of the divide between quantum and classical is by appeal to the conscious experience of the observer: her "seeing it" precipitates the collapse. As will be evident, this also agrees with the Copenhagen interpretation of the delayed choice experiments discussed above, according to which the observer brings phenomena into being.

To Einstein and Schrödinger this was anathema. For them physics is about the nature of reality, not about what the community of experimental physicists decide to do, and certainly not anything to do with the supposed powers of conscious reflection

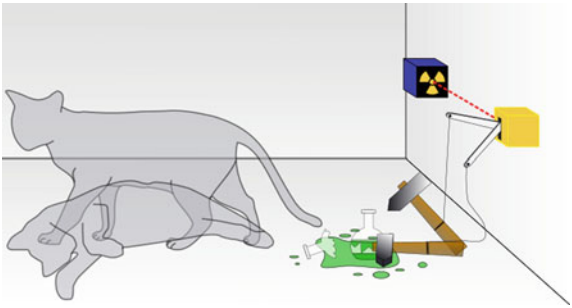


Fig. 8.3 Schrödinger's Cat (Illustration taken from the Wikipedia article, "Schrödinger's Cat", https://en.wikipedia.org/wiki/Schrödinger%27s_cat, accessed January 27, 2019.)

to bring things into being.[18] To illustrate the absurdity of this, Schrödinger sardonically observed, "one can even set up quite ridiculous cases" (Schrödinger 1935b, 157), and proposed his "cat paradox". In this thought experiment a cat is shut up in a steel chamber "along with the following diabolical device": a Geiger counter containing a tiny amount of radioactive material which has an equal probability of decaying or not decaying in the course of one hour. A decay would trigger the Geiger counter and, through a relay to a hammer that would shatter a phial of hydrocyanic acid, poison the unfortunate animal; if after an hour, there had been no such decay, the cat would still be alive (Fig. 8.3).

"The ψ-function of the entire system," Schrödinger wrote, "would express this by having in it the living and dead cat (pardon the expression) mixed or smeared out in equal parts" (157). Clearly, no such "smeared" cat is ever observed: we only ever observe a cat that is alive or dead, not one in a state that is a superposition of being alive and dead.

Schrödinger's Cat Paradox has acquired a great notoriety, despite the fact that it was tossed off in a short paragraph in the course of a long and detailed analysis. What this "ridiculous case" is supposed to bring into relief is the absurdity of the dogma that there is no actual state of affairs before the intervention of the observer (or her consciousness). As Einstein remarked to his friend in approval, "Nobody really doubts that the presence or absence of the cat is something independent of the act of observation" (Przibram 1986, 39). This is not an attack on the very idea of entanglement, as is sometimes supposed. Nor is Schrödinger here opposing what we might call Bohr's "sober" view, according to which a measurement does not actually require a physicist or knowing subject. All that is required is that the quantum system interact with its environment so as to produce some observable effect—in what Wheeler termed an "irreversible act of amplification".[19] For instance, when an alpha particle is emitted by a nucleus, it is described by a "spherical wave that

---

[18] Cf. Schrödinger's sardonic remark: "It is rather discomforting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter's mercy in spite of his having no access to it" (Schrödinger 1935a, 556).

[19] Cf. also John Stuart Bell: "The only 'observer' which is essential in orthodox practical quantum mechanics is the inanimate apparatus which amplifies microscopic events to macroscopic consequences" (Bell 1987, 170).

continuously emanates in all directions from the nucleus" (Schrödinger 1935b, 156), but the screen only "lights up at *one* instant at *one* spot" at a time (157). It is therefore "naïve" to suppose that this wave function represents a "blurred" reality. What is typical in such cases, Schrödinger remarks, is that the indeterminacy that applies to the quantum domain is "transformed into a macroscopic indeterminacy, which can be *resolved* by direct observation" (157). The question is how? and when? If the results of measurements—or indeed, any in principle observable events at all—are always perfectly definite, and this involves the "collapse of the wave function", then how and when does that collapse occur? As Bell has described the problem:

> The continuing dispute about quantum measurement theory … is between people who view with different degrees of concern or complacency the following fact: so long as the wave packet reduction is an essential component, and so long as we do not know exactly when and how it takes over from the Schrödinger equation, we do not have an exact and unambiguous formulation of our most fundamental physical theory. (Bell 1987, 51)

Of course, this presupposes that there must actually be a collapse of the wave-packet, something Bell himself, among others, is inclined to reject. Carl Hoefer has summarized the kinds of objections to collapse that are common to adherents of both the MWI and de Broglie-Bohm theories:

> It has the virtue of solving certain problems such as the infamous Schrödinger's cat paradox, but few philosophers or physicists can take it very seriously unless they are instrumentalists about the theory. The reason is simple: the collapse process is not physically well-defined, is characterised in terms of an anthropomorphic notion (*measurement*), and feels too *ad hoc* to be a fundamental part of nature's laws. (Hoefer 2016, §4.4)

But before we turn to interpretations that reject collapse, let us be sure that we understand the problem.

As we have seen, Schrödinger rejected any anthropomorphic notion of measurement, and, like "the sober Bohr", interpreted it as involving an irreversible act of amplification. He did not regard wave function collapse as solving the cat paradox, but proposed the latter to highlight the measurement problem. The problem is not that the theory predicts discrete events, whereas the evolution of the wave function is linear and continuous in time. *That* problem is already resolved by the Born interpretation, according to which eigenvalues corresponding to the discrete events in question are predicted with probability amplitudes given by the coefficients of the eigenstates in the appropriate superposition.[20] If there is a lack of determinism in the theory regarding the prediction of particular events (such as where and when an alpha particle is emitted by a nucleus), this is due to the functioning of ψ as a probability amplitude, not to von Neumann's assumption that there is a discrete, process of collapse caused by the observer. It was this indeterminism that underlay the Bondi-Reichenbach-Ellis conception of becoming described above.

---

[20] I therefore disagree with formulations of the measurement problem such as this: "In short, the measurement problem is this: Quantum theory implies that measurements typically fail to have outcomes of the sort the theory was created to explain" (Goldstein 2017). But any probabilistic theory fails to predict outcomes in this sense: a probability of ½ for a coin toss yielding heads does not predict whether it will actually land on heads or tails.

In contradiction to this, it has been maintained (for example, by Hoefer) that quantum theory is perfectly deterministic. If the essence of quantum theory is evolution of a system governed by the Schrödinger equation, then once the state function is determined from the initial conditions, all future states are determined.[21] The projection postulate, on this view, introduces an unwarranted and unpredictable indeterminacy into an otherwise perfectly deterministic theory, and should therefore be rejected. This presumes, however, that what must be determined in the theory are the future states of the system corresponding to a given ψ-function. But on the Born interpretation, a state is not the state of a system in the classical sense—this is the naïve view rejected by Schrödinger—but a sort of "expectation-catalogue", to use his term (Schrödinger 1935b, 161–2), allowing us to predict events with certain probabilities, assuming the same situation should recur. "Causal determinism", I submit, concerns the determining of *events*, not the states used to predict their probabilities. In this sense the theory is indeterministic, and the de Broglie-Bohm theory is no less so, since it makes the same predictions.[22] I will return to this point below.

What, then, is the measurement problem? Several problems may be identified, and the following is likely not an exhaustive list. First, there is the problem of the alleged divide between the quantum realm and the classical—often rather inaccurately described as that between the microscopic and the macroscopic.[23] Given quantum entanglement, how is it even possible that perfectly unambiguous and macroscopically observable events can occur? Second, if collapse is not a physical process, how is it to be understood? Third, if there is no collapse, is entanglement a global phenomenon, and are we committed to there being a state-function of the whole universe?

The first of these questions has now basically been resolved to the satisfaction of most interpreters. It depends on the notion of *decoherence*. The basic intuition was developed by David Bohm in his (1952) to account for the appearance of wave function collapse, a notion that was exploited by Hugh Everett in developing his "relative state" interpretation (Everett 1957). It was then given a precise rendering by H. Dieter Zeh in 1970, and has since been subjected to further systematic elaboration and development. The key to decoherence is this: the Schrödinger equation is applicable only to a closed system, but typical macroscopic states are never isolated from their environments. An isolated system evolving in time according to the Schrödinger equation will be such that the phases of the states of its subsystems are all coherent with one another. Such coherence is what is responsible for the entanglement of

---

[21] Cf. Carl Hoefer: "The fundamental law at the heart of non-relativistic QM is the Schrödinger equation. The evolution of a wavefunction describing a physical system under this equation is normally taken to be perfectly deterministic" (Hoefer 2016, Sect. 4.4). For an extended treatment of the claim, see Earman's *A Primer on Determinism.*

[22] Cf. Bohm (1957, 116): "Finally, let us note that in our model we have not insisted on a purely causal theory, for we have also utilized the assumption of random fluctuations originating at a deeper level."

[23] Lasers and super-cooled fluids are both macroscopically observable quantum objects. Zurek gives the example of the Weber bar (a gravity-wave detector), which "must be treated as a quantum harmonic oscillator even though it may weigh a ton" (Zurek 2002, 4).

these states, and this is what is exploited in quantum computing. Complete isolation is, however, difficult to maintain, so that when a quantum system is in interaction with its environment, this coherence is lost to the system. It is “leaked out” into the environment in the form of correlations with the states of its component systems. The result is that the “off-diagonal terms” representing interference become minimized, and the resulting probabilities add like classical probabilities, representing both system and apparatus in definite though unknown states, with the predicted degree of probability. This is known as quantum decoherence, while the selection induced by the environment of states that do not cohere with one another is known in the trade as *einselection*.[24]

This may be conceived as follows. When a previously isolated quantum system enters a complex environment, it forms superpositions with elements of that environment in rapid succession. Each of these superpositions involves correlations of the phases of the interacting systems. But in a typical macroscopic environment, there is an enormous number of these, so that the configuration space in which the combined system is represented increases in volume in a flash. It may help to explain this by reference to the example of α-particle emission explained by George Gamow. As Nevill Mott already noted in his (1929), there is an apparent difficulty here, in that the wave function of the particle is represented by a spherical wave, but “the α-particle, once emerged, has particle-like properties, the most striking being the ray tracks that it forms in a Wilson cloud chamber” (Mott 1929, 129). But as Mott explains, the discrepancy is a result of “our tendency to picture the wave as existing in ordinary three-dimensional space, whereas we are really dealing with wave functions in the multispace formed by the co-ordinates both of the α-particle and of every atom in the Wilson chamber” (129). Taking only two atoms, he shows that “they cannot both be ionized unless they lie in a straight line with the radioactive nucleus” (130). But there are some 10^{24} atoms in a typical cloud chamber. So the significance of Mott’s result is that once one atom has been ionized, there will be an extremely high probability of ionization of atoms lying in an approximately straight line from the nucleus. In the gigantic configuration space of all these atoms, there will therefore be histories of diverging paths through this space representing such rays, completely swamping the effects of the interference between the elements of the original system. These roughly straight trajectories will therefore be einselected. As Zurek explains, such einselection of effectively classical trajectories “enables one to draw an effective border between the quantum and the classical in straightforward terms, which do not appeal to the ‘collapse of the wave packet’ or any other such *deus ex machina*.” (Zurek 2002, 6).

Now let us turn to the interpretation of collapse. On von Neumann’s interpretation it involves projection onto an eigenstate, with projection conceived as a physical process that is “discontinuous, non-causal and instantaneously acting” (von Neumann 1932, 551). One of the main reasons for supposing such a process in addition to Schrödinger evolution was to account for how unambiguous (“classical”) values of

---

[24] Here “ein” stands for “environmentally induced superselection” (with perhaps a hint of reference to Einstein, as though Albert were induced by his environment to select a super stein of lager?).

physical quantities could be obtained on “measurement”. But as we have seen, decoherence explains that. So if the appearance of classical values is all that is desired, then it appears that the projection postulate—the postulate of a separate type of discontinuous, physical process—is unnecessary. Given this, together with the apparent vagueness about how and when such a process is supposed to take place, and its seeming extraneousness to the axioms of quantum theory, Bohm advocated rejecting it altogether. As we shall see, Everett, together with perhaps the preponderance of theoretical cosmologists, have followed him.

On Everett’s interpretation (I shall defer discussion of the de B-B interpretation to the next section), there is never any collapse, only the appearance of collapse to an observer who is necessarily constrained to be in one branch of the superposition. In this branch, the observer’s wave function is coupled with that of its environment (with that of the Geiger counter and amplifier, in the case of the cat) in such a way that when the observation is made the cat is found dead, say, while on another branch an up-to-that-moment identical observer observes the cat alive and well. There is, on his theory, no transition from ‘possible’ to ‘actual’, “nor is such a transition necessary for the theory to be in accord with experience” (Everett 1983, 320).

One obvious worry about the MWI concerns probabilities. The above scenario presumes only two, equiprobable, discrete outcomes of an observation. If the radioactive sample were bigger or left longer, the cat might have had a 90% chance of being killed. If that means 90% of the worlds are ones in which the cat dies, how could this be subjected to test? For even though quantum theory applies to single systems (like Wheeler’s photon arriving from a quasar), “it is a statistical theory, so it must be tested on ensembles of such systems. Now, it does not seem likely that anyone will ever test a theory on an ensemble of worlds” (Torretti 1999, 337).[25]

But I think there is a related objection that is even more profound, one that will also infect the theories of quantum gravity I shall be considering later: if all possible events are actual, what sense does probability even make?[26] On the MWI, there are infinitely many branches, each corresponding to a different possible outcome of a measurement by an observer, and necessitating the existence of a universe corresponding to that branch. The wave function of the whole universe (more accurately, the multiverse) is a stupendously complex entanglement of all these branching worlds. But there is no collapse, just an evolution of relative states. So, as long as these are thought of as possibilities with their own probabilities, there are, in fact, no events. Consequently

---

[25] Cf. Maudlin: “since there are no frequencies in the theory, there is nothing for the numerical predictions of the theory to mean” (Maudlin 2002, 5). (He uses an almost identical “cat scenario”, where the probability of one result is 99%.) Lee Smolin has made two attempts to circumvent this objection: (1) his cosmological natural selection hypothesis, according to which there are ensembles of universes that “reproduce when black hole singularities bounce to become new regions of spacetime” (Unger and Smolin 2015, 455); and (2) his “real ensemble interpretation” (489), according to which quantum mechanical ensembles consist in different copies of a microscopic system that interact with each other by copying values of the ‘beables’ of the other members of the ensemble (Smolin 2011, 3).

[26] There have, of course, been many attempts to answer that question, as Steve Savitt reminds me. See in particular David Wallace’s *The Emergent Multiverse*, (2012), Part II.

Everett postulates that they are all actual, not merely possible. The events we see are all these possibilities-become-actualities along the branch we are on.

But what are these events? According to Everett's original interpretation, they are all the possible outcomes of any possible measurement. So in the case of the decay of the alpha particle in Schrödinger's thought experiment, since this could decay at any time during the hour, the universe must split into branches corresponding to these different states of affairs of the particle being observed to decay at every single instant of the hour. But the number of instants in an hour (according to set theory) is non-denumerable.[27] Worse (if that's possible), there is no unique set of eigenstates into which a given wave function can be decomposed (no preferred basis). If all correlated states of a combined system, such as observer and the quantum system being observed, are equally actual, then all possible decompositions of the wave function into different sets of eigenstates would require whole actualized worlds in which they occurred.

The no-preferred-basis objection can be parried by appeal to decoherence. The only possible decompositions, on this view, would be those that have been einselected. For example, in the case of Schrödinger's cat, the probabilities corresponding to the cat's being observed to be alive and its being observed to be dead (after the hour has passed) are both arbitrarily close to one half, and the probability of its being observed in a superposition of the two has been reduced to near zero. But these are *probabilities of events*, not events. This, I believe, illustrates an ambiguity in the whole discussion of wave function collapse. In the Copenhagen Interpretation, collapse was interpreted as a discontinuous transition induced by an observer performing a measurement, requiring the introduction of a separate kind of physical process (as in von Neumann's projection postulate). But without this postulate, the theory still predicts certain values or outcomes if interactions take place, by application of the Born Rule. If two quantal systems are suitably isolated from their surroundings, but in a state of interaction with one another, they remain in a state of superposition. Now, we can only verify this fact by setting up several such identical experimental arrangements breaking this isolation, and analysing the statistics of the events that occur. That such events do occur (with their statistics indicating entanglement of their states) is arranged by having them subject to a macroscopic determination, in what Bohr has called an irreversible act of amplification. The events must occur, otherwise we will not be able to test the theory. If they do occur, we reassign wave functions to accord with our new knowledge of the situation, if we want to make further predictions. But this does not entail that the events are caused to occur by a discontinuous physical process, as von Neumann proposed. The apparent "jump" from the original state function to the eigenstate is, as Torretti insists, "the transition from an initial situation in which a chance event is being expected to a final solution in which the outcome of chance is already given" (Torretti 1999, 341). The fact that the theory only predicts whether (among the events selected out by the experimen-

---

[27] If I correctly understand Rovelli's *relative state interpretation* of quantum theory, a similar criticism must apply; although here there are not eigenstates (quantum events) in an absolute sense, but only relative to other states.

tal arrangement) a certain event will occur with a certain probability is simply to say that the theory is stochastic.[28] It is stochastically deterministic, but not casually deterministic. But there must be events (and must have been before there were any physicists to observe them): states are not directly observable, they are only inferred from events.[29]

This brings us to the third aspect of the measurement problem mentioned earlier. On the accepted analysis of decoherence explained above, quantum entanglement is not eliminated, just made effectively unobservable to an observer on one branch. On such a view, then, all interactions since the origins of the universe have entangled the systems in question. Accordingly, there should be one wave function for the whole entangled universe. For cosmologists, this is a major motivation for accepting the MWI.

But this is a very abstract point of view, and, I submit, fails to take into account the conditions under which we apply quantum theory. First, the wave function must be determined, for a given system, in a given environment. Usually this environment is modelled by classical potentials, and then adjusted accordingly. (This is analogous to GR, where the mass-energy tensor is formulated by classical modelling.) We abstract from the possibly infinite complexities of the environment, and treat the system + apparatus as effectively isolated during the course of the analysis. But as Bohm and Bell have both stressed, we do not know all the initial conditions. It is hubris to suppose that the wave function contains all the knowable information when physical features of the environment that we need to know in order to formulate the wave equation could be infinitely complex and not contained in the way we have modelled the environment. Second, the application of quantum theory is irreducibly local. We can calculate the wave function of an alpha particle that has been emitted by a nucleus, and treat it as a free particle. But if it then interacts with its surroundings (for example with a Geiger counter) then its wave function will become entangled with the wave function characterizing those surroundings, and to make further predictions one would apply quantum theory by beginning with the wave function for the combined,

---

[28] An alternative interpretation of this stochasticity as a real spontaneous collapse is the GRW approach pioneered by GianCarlo Ghirardi, Alberto Rimini and Tullio Weber; see Ghirardi (2018). See also Dorato and Esfeld (2010), who argue that GRW "is a serious candidate for being a fundamental physical theory" (2010, 41), and that it resolves various problems of interpretation in quantum theory in a philosophically satisfying way.

[29] Cf. Jammer (1974, 474): "If the whole physical universe were composed only of microphysical entities, … it would be a universe of evolving potentialities (time-dependent ψ-functions) but not of real events". For a thorough analysis of the measurement problem in full, formal detail see Jeffrey Bub's *Interpreting the Quantum World*. He argues that the appeal to einselection on the MWI view to give a resulting "mixture with respect to events associated with the pointer basis, not only fails to account for the occurrence of just one of those events, but is actually inconsistent with such an occurrence (considering the origin of this mixture, and assuming the orthodox interpretation principle)" (Bub 1997, 231). In his Chap. 4 he proves a theorem to show how to construct 'no-collapse' interpretations that account for the existence of facts (events), by giving up the orthodox Dirac-von Neumann "'eigenvalue-eigenstate link' in favour of an interpretation that takes an appropriate preferred observable as having a determinate value independently of the quantum state" (232). Torretti (1999, 359) advocates a similar position.

entangled system. If its future encounters are contingent, as they certainly should be if the intervening factors (such as times of decay, etc.) are subject to quantum indeterminacy, then all assignments of wave functions are provisional and contextual. But (even leaving aside the problem of initial conditions) the universe is not local, has no environment, and there is no external time with respect to which its development could be taken to be progressing. From this perspective, it is not at all obvious that there should be a wave function for the entire universe, however convenient that might seem to cosmologists.

To sum up: it is not a shortcoming of quantum theory that it does not predict precisely when and where a given event should take place. That is the objective chance implicit in the Born Rule. In purporting to solve the measurement problem by reducing reality to relative states, the MWI interprets all possibilities as actual events in some given universe, so that in any given universe such as ours, all events are actual. In so doing it eliminates not just the interpretation of wave function collapse as a physical process, but objective chance itself. By interpreting the evolution of states representing probabilities of events as events, it obscures the very meaning of "event".

But a more direct attack on the idea that quantum theory involves irreducible chance is made in the de Broglie-Bohm approach, to which I shall now turn.

8.4 Non-locality and Relativistic Quantum Mechanics

Schrödinger came up with his cat paradox as a variant on a similar thought experiment Einstein had mentioned in their correspondence (where the trigger ignites an explosion of dynamite rather than breaking a phial to poison a cat).[30] Einstein had also preceded him in print with a thought experiment designed to undermine the Copenhagen Interpretation. This was the paper co-written with Boris Podolsky and Nathan Rosen, published earlier in the same year, containing the famous *EPR paradox*. This paradox also exploits the continued entanglement of systems that have once interacted. In the example they use, the systems are assumed to have interacted between $t = 0$ and $t = T$, but to have had no further interaction afterwards. If the momentum of the first is measured, this collapses the wave function into an eigenstate of momentum. Because of entanglement, the second system will then be known also to be in an eigenstate of momentum. On the other hand, if the experimenter had chosen to measure the position of the first system, then the second system would be in an eigenstate of position. But it could not then be in an eigenstate of momentum, since position and momentum are canonically conjugate. "We therefore see that, as a consequence of two different measurements performed upon the first system, the second system may be left in states with two different wave functions" (Einstein et al. 1935, 140). But according to the authors' "criterion of reality",

---

[30] See Jammer (1974) for a discussion and references.

> If, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity. (138)

Since, by hypothesis, the second system was not disturbed by either measurement on the first, and it could have been in a definite state of position or momentum depending on what experiment was performed on the first, then (they inferred) it must have had a real value of position or momentum independently of measurements on the first system. But according to quantum theory, it could not have been known with certainty to be in a definite state of position and momentum simultaneously, since these are canonically conjugate quantities represented by non-commuting operators. It follows that the wave function does not give a complete description of reality.

In his reply Bohr (1935) insisted that the wave function does not describe individual systems, but the "whole experimental arrangement" of system plus apparatus. So the two alternative experiments on the first system are on two different states of affairs. Thus one cannot infer that the same second system is involved in both cases, and there is no contradiction forthcoming. Although Bohr's reply was by no means easy to understand in detail, the great majority of physicists regarded it as a successful rebuttal to Einstein's thought experiment. But the gist of Bohr's identification of the flaw in the EPR reasoning was clarified by David Bohm in his 1951 book, *Quantum Theory*. Two non-commuting observables, such as the position and momentum of an electron, cannot simultaneously exist with precisely defined values, given the Indeterminacy Principle. Rather, either of them "is potentially capable of becoming better defined at the expense of the degree of definition of the other, in interaction with a suitable measuring apparatus" (Bohm 1951, 365). Since "the realization of these potentialities depends just as much on the systems with which it interacts as on the electron itself," however, "there are actually no precisely defined 'elements of reality' belonging to the electron" such as EPR had supposed.

Bohm had explained the fallacy of EPR's reasoning using a reformulation of the thought experiment (now known as EPRB) in terms of spin. In this scenario, a molecule in a state of spin 0 disintegrates, in such a way that it spin is unaffected, into two atoms, **A** and **B**, whose spin values in a given direction are either $+\hbar/2$ or $-\hbar/2$ (Bohm 1951, 359). Because spin is conserved, these values must always be opposite in sign in order to add to 0. The spin can then be determined by having the atoms traverse a magnetic field, as in the Stern-Gerlach experiment, in which they will be deflected either up (+ve spin) or down (−ve spin). So if atom **A** is found to have a spin component of $+\hbar/2$ in the x-direction, say, then it can be "predicted with certainty" that atom **B** will be found to have spin $-\hbar/2$ in the x-direction, and vice versa. So by the EPR criterion, **B** must really have such a value of spin in that direction. The rub is this: the spin components in orthogonal directions are represented by non-commuting linear operators. So if we know with certainty that **B** has spin $-\hbar/2$ in the x-direction, then we are precluded from assigning it a definite value of $+\hbar/2$ or $-\hbar/2$ in the y- or z-direction. But these are the only spin values it can have in those directions. We could equally well have measured **A**'s spin in the y- or z-direction, and then we would be able to predict with certainty **B**'s spin value in either direction. The EPR conclusion for this scenario would then be that there are elements of reality concerning **B** that the theory cannot predict. But according to quantum theory, precisely in the case where it is possible to predict **B**'s spin value in the x-direction (because we have measured **A**'s spin in this direction) it is *not* possible to "predict with certainty without in any way disturbing the system" the components of **B**'s spin in the y- or z-directions.

Still, as Bohm subsequently realized, this refutation of EPR presupposes that the probabilities predicted by quantum theory are irreducible, and could not in principle be the result of a shortcoming of the theory. It supposes that "the indeterminacy relationship should have an absolute and final validity" (Bohm 1957, 84). Under the stimulus of discussions with Einstein the year after his book came out, however, he was able to devise a theory (Bohm 1952) in which the indeterminacies were not fundamental, but a result of statistical fluctuations in "hidden variables" on a sub-quantum level. On this proposal, particles always have well-defined (but unpredictable) trajectories, and Schrödinger's ψ-function is a guiding wave—just as on de Broglie's "pilot wave" theory (de Broglie 1927), as Bohm "later learned" (Bohm 1957, 110). De Broglie had abandoned his theory in the face of harsh criticism, and in the 1950s it was still believed that "hidden variable" theories of this kind had been proved to be impossible.[31] But Bohm realized that such impossibility proofs must have begged the question, since his and de Broglie's theory posited particle trajectories at a subquantum level, yet gave the same statistics as the orthodox theory.

The basic idea of Bohm's original proposal is mathematically straightforward. Schrödinger's ψ-function is represented as the product of two factors, $\psi = R\exp(iS/\hbar))$, where R and S are real-valued functions. On substituting ψ into the Schrödinger equation, two equations are derived, each linking R and S. R can be interpreted as the square root of the particle density, and by analogy with the classical case, the divergence of $S(x)$ can be interpreted as giving the particle's momentum at the point x. Also by analogy with the classical case, Bohm was able to define a quantum mechanical potential as a function of R.[32] But the precise form of the force is unimportant, and subsequent versions of "Bohmian mechanics" have been given in which no quantum potential is derived. What is essential, though, is that "the force is such as to produce a tendency to pull the body into a region where $|\psi|$ is largest" (Bohm 1957, 112). Thus in this theory ψ is "regarded as a mathematical representation of an objectively real force field" (Bohm 1952, 376). If at some initial time t_0 the configuration of the system is a random one given by the Born rule, then

---

[31] Von Neumann had given a proof of the impossibility of a hidden variables theory in his (1932). The received view circa 1950 is well summarized by Born: "No concealed parameters can be introduced with the help of which the indeterministic description could be transformed into a deterministic one. Hence if a future theory should be deterministic, it cannot be a modification of the present one but must be essentially different" (Born 1949: 109).

[32] The quantum mechanical potential is given by $U(x) = -\hbar^2/2m\{\nabla^2 R/R\}$, and this together with the classical potential would determine the exact position of the particle, assuming the initial conditions were known.

it is a consequence of the Schrödinger equation that it will always accord with that rule, provided the system remains isolated.[33]

This theory allowed Bohm to give realistic interpretations of the two-slit and delayed choice experiments, as we saw in the previous section. The insertion of the half-silvered mirror is a change in the physical environment that results in changes in the force field: the possible trajectories of the photons are instantaneously altered by the action of the quantum potential. "The statistical tendency to appear where $|\psi|^2$ is greatest is due to the effects of the 'quantum-force' while the random motions explain why the precise points at which the various particles appear fluctuate in an irregular way" (Bohm 1957, 115). It also enabled him to explain the EPRB thought experiment in a similar way. The correlation of atom **B**'s spin-direction with **A**'s is the result of the measurement on **A** changing the force field experienced by **B** in such a way as to orient the spins in opposite directions along the same axis, whichever axis is chosen for the magnets in the experiment. (Note that this explanation supports Bohr's contention that it is the "whole experimental arrangement" that is being discussed, not isolated particles. This context-dependence of the wave function is now known in the jargon as *contextuality*.) A further major appeal of Bohmian mechanics is that there is no "measurement problem". Particles always have well defined positions, so there is no mystery that these are manifested in experiments.

There is one troubling feature of Bohm's theory, though, that I have left unmentioned so far. This is that the action of the quantum potential in transmitting momentum among the particles, assumed to be moving randomly at the sub-quantum level, is *instantaneous*.[34] This is clearly contrary to the special theory of relativity, if this is understood as prohibiting any processes that occur faster than *c*, as I argued in Chap. 6. But because the fluctuations are uncontrollable (though correlated), Bohm contends that no signal can be sent, so that no contradiction with relativity can arise.[35] We will return to this point shortly.

Now, it might be thought that this correlation of spin-axes is not particularly mysterious. John Bell ([1981b] 1987) wittily compares it to the case of Dr. Bertlmann, who always wore socks of different colours. So if we see a pink sock appearing as he walks into the room, we can be sure the second sock is not pink. Similarly, why not suppose that the spin-orientations of the atoms in the EPRB experiment simply have precise values that are randomly selected, but always correlated, like Bertlmann's sock colours? What Bell showed by an ingenious and simple mathematical argument was that such an assumption (the "elements of reality" assumption made by EPR, coupled with the assumption that there is no influence going from the detection of one "sock" to the other) leads to statistical predictions that contradict those of quantum

---

[33] This is known as the "quantum equilibrium" hypothesis. Valentini (2002) has shown that "instantaneous signals at the statistical level" are possible if (and only if) this initial equilibrium condition is violated for an isolated system.

[34] Bohm (1952, 389): "Thus, the 'quantum-mechanical' forces may be said to transmit uncontrollable disturbances instantaneously from one particle to another through the medium of the ψ-field."

[35] "The reason why no contradictions with relativity arise in our interpretation despite the instantaneous transmission of momentum between particles is that no signal can be carried in this way." (Bohm 1952, 390).

mechanics. This result (or a more generalized form of the argument) is known as Bell's Theorem.

If we assume that an experiment on one atom has no influence on the result when we measure the other (as with Bertlmann's socks), then the correlations must be determined in advance by the orientation of the magnets and the state of the source. Indeed, if we measure the spins with the magnets oriented along the same axis the spin values are always opposite, whereas if the orientations of the magnets are chosen at random, then (when the experiment is repeated) the spins agree or disagree precisely half the time, just as we would expect with Bertlmann's socks. But now if the magnets are given orientations that are not parallel, under the same assumption of their correlations being determined in advance, the statistics produced will be incompatible with those predicted by quantum theory (and verified by experiment). With the orientation of the second magnet swung through 60° (so that the axes are aligned at 120° apart instead of 180°), quantum theory predicts that the magnets will record the same spin three-quarters of the time, and the opposite spin one quarter of the time. On the assumption that the correlations result only from being determined in advance, it is not possible that such a large measure of correlations could occur. This is what Bell demonstrated in his famous theorem by deriving an inequality based on the "Bertlmann's socks" assumptions of existing "elements of reality" and no influence from one detector on the other, an inequality that is violated both by what quantum theory predicts, and by experimental results.

Crucially, the detections made in each correlated pair of results in the two wings of the experiment can be arranged so that they are spacelike separated. Then even if a light signal were sent out when a detection was made in the first wing, it would arrive too late to influence the second. This is exactly what was done in a series of experiments conducted by Alain Aspect and co-workers.[36] From an analysis of such scenarios, Bell inferred that this is only explicable in terms of superluminal influences passing from one side of the experimental apparatus to the other:

> The EPRB correlations are such that the result of the experiment on one side immediately foretells that on the other, whenever the analyzers happen to be parallel. If we do not accept the intervention on one side as a causal influence on the other, we seem obliged to admit that the results on both sides are determined in advance anyway, independently of the intervention on the other side, by signals from the source and by the local magnet setting. But this has implications for non-parallel settings which conflict with those of quantum mechanics. So we *cannot* dismiss intervention on one side as a causal influence on the other.[37]

Now there is a vast and complex literature on Bell's Theorem, and this is not the place to go into the various elaborations and distinctions that have been made. As

---

[36] Cf. Maudlin (2002, 24): "Aspect's experiment was so contrived that the setting of the equipment at one side could not be communicated, even by light, in time to influence the other side".

[37] (Bell [1981] 1987, 149). Cf. his conclusion in his 1964 paper: "there must be a mechanism whereby the setting of one measuring device can influence the reading of another instrument, however remote. Moreover the signal involved must propagate instantaneously, so that such a theory could not be Lorentz invariant." (Bell [1964] 1987, 20). Cf similar remarks about the violation of Lorentz invariance at (Bell 1987, 20 and 154).

promised, I will restrict comments to what seems relevant to time and the reality of becoming.

The first point I wish to draw attention to is the oddness of Bell's conclusion that we must admit that intervention on one side of the apparatus has a "causal influence" on the other. As we have seen, this is precisely what Bohm denied in his analysis of EPRB, assuming his hidden variable theory: although intervention on one side causes an instantaneous change in momentum of the supposed particle trajectories, and one may, I suppose, describe this as a "causal influence", there is no known causal mechanism which would reproduce the statistical results. Because of the unpredictable fluctuations undergone by the particles at the subquantum level, "no signal can be carried in this way" (Bohm 1952, 390). The larger-than-expected correlations show up in the statistics, but not in any given run of the experiment. This is jocularly summarized in the aphorism "there is no such thing as a Bell telephone", namely a means of exploiting the violation of the Bell inequalities for superluminal signalling. Nevertheless, Bell asserts that an instantaneous propagation of a signal from the source is the only way to explain the violations of his inequality, and that such non-locality is "characteristic … of any such theory which reproduces exactly the quantum mechanical predictions" (Bell [1964] 1987, 14). "It now seems", he argued in 1981, "that the non-locality is deeply rooted in quantum mechanics itself and will persist in any completion" (Bell 1987, 132). He also suggested that "it may well be that a relativistic version of the theory, while Lorentz invariant and local at the observational level, may be necessarily non-local and with a preferred frame (or aether) at the fundamental level" (133).

Second, there is the oddness that a result proved on the basis of Galilean-relativistic QT is offered as an argument for doubting the universal validity of Lorentz invariance. One might think that this is analogous to rejecting the principle of retarded action that is implicit in special relativity on the grounds that it is contradicted by the Newtonian theory of gravitation, where the gravitational potential is also assumed to be able to act instantaneously.

In assessing Bell's claims, I think here we need to make a firm distinction between two different cases: (1) that of a hidden variables theory where the positions of the particles are precise at all times, even though unpredictable, and (2) that of orthodox quantum theory, where no precisely defined positions can be assigned to a quantum system until it is has interacted with something equivalent to a suitable measuring apparatus, as in Bohm's original response to EPR (Bohm 1951, 365). In the first case, assuming a de Broglie-Bohm-type hidden variables theory, the correlations are explained by an instantaneous rejigging of the possible particle trajectories, and the EPRB statistics are explained by the fact that the particle positions are constrained to agree on average with the predictions given by the wave function. But this constraining is effected by the instantaneous action of the quantum potential (or some equivalent action of ψ interpreted as representing an objectively real force field), thus violating Lorentz invariance. It is important to see that this violation does not consist merely in this rejigging taking place in a preferred frame of reference.[38] If my arguments in Chap. 6 are correct, the idea of any physical *process* taking place instantaneously is incompatible with the assumption of retarded action, which is one of the constitutive assumptions of the Lorentz transformations as a group. So, given Lorentz invariance, there can be no physical process that effects the changes that the de Broglie-Bohm theory purports to explain.

Moreover, the very idea of precise particle positions is incompatible with Lorentz invariance at a very deep level. If we assume a Galilean-invariant quantum theory, the spatial location operator allows one to calculate expectation values of the position q of a quanton in a given environment at a time t. This is a dynamical variable specific to that system; it is not the x in $\psi(x)$, which labels a point in the abstract configuration space, this configuration space being isomorphic to ordinary 3-space only in the case of a single quanton. On the de Broglie-Bohm interpretation, expectation values for q correspond to average values over precise particle locations, a point in the configuration space encoding the positions of all the constituent particles. In a relativistic quantum theory, on the other hand, there are no Lorentz-invariant location operators (either spatial or temporal). So there is no longer any possibility of interpreting expectation values as average values over precise particle locations at a time, since there simply are no expectation values for particle locations that are Lorentz-invariant. Furthermore, it is provable from basic quantum principles together with Lorentz-invariance that particle number is not conserved: there is no such a thing as the number of particles at any time. Consequently, there is no viable relativistic theory of particles with precise locations such as is presumed in the Galilean-invariant de Broglie-Bohm interpretation. Hence the need for quantum field theory, where particles are reinterpreted as excitations of fields.

But what if, (2), we do not assume a hidden variables theory where the positions of the particles are precise at all times? Now there is no question of an uncontrollable disturbance passing from a particle in one precise position to another particle in another such position, since there are no such precise positions until they have been bought about by the irreversible interaction of a quanton with something that would detect it, such as a Geiger counter or screen. Still, given the contextuality noted earlier, the potential for a given event to occur depends on the whole experimental set-up; and this can be arranged, as in Aspect's experiments, so that the events of detection in the two wings of the experiment are spacelike separated. The question is, don't the resulting correlations demand an explanation in terms of a non-local influence, as Bell alleges?

Here I think the following comments are in order. First, it is not at all clear what such a non-local influence could consist in. Even putting aside the above objection that it could not be a physical process, if it were a process it would be a very strange one. First, it would be unattenuated. As Maudlin writes, "If Aspect had put one wing of his experiment on the moon, he would have obtained precisely the same results"

---

[38] In fact, as Kent Peacock has insisted, there being a preferred frame of reference is not in itself a violation of Lorentz invariance, "any more than the fact that any subliminal motion is associated with a distinguished frame (namely the rest-frame of the system)." (Peacock 2018, 120).

(Maudlin 2002, 22). Second, it would be weirdly selective, coming into play, not when the magnets were aligned along a straight line or perpendicularly, but only when they had some intermediate angle of relative orientation.[39] Most importantly, though, it is not the individual correlations in any one run of the experiment that need explaining, nor the sequences of results on either wing, each of which sequences taken in isolation is entirely random. The correlations lie in the statistics, not in individual event pairs. Finally, the correlations are tracked in configuration space. But these are correlations among predicted values, correlations among propensities for events to occur, not correlations among events. So an interpretation of them as "influences", as if they are physical processes, does not seem justified.

So what are we to conclude? Events occur, and when they do so, the connections among them fall along the trajectories described in relativity theory. But provided quantum systems remain isolated, the probabilities of events evolve in such a way that correlations among them occur even across spacelike intervals—at least statistically, even if not individually. This is highly counterintuitive, but it does not seem to constitute grounds for giving up Lorentz-invariance.[40] One can regard quantum probabilities as representing tendencies to manifest or actualize, tendencies that are indeterministic. But whatever the links among such tendencies, their actualizing is always local. On this picture, a local version of the Bondi-Reichenbach-Ellis conception of becoming as "becoming-determined" seems vindicated: becoming is simply the local actualization of tendencies.[41]

8.5 Spacetime and the Quantum

General relativity, for all its successes in cosmology, is from a quantum perspective still a classical theory: it presumes precise trajectories for particles of dust moving along its geodesics, and it fails to take into account that in regions of spacetime where matter is very dense, there will be a corresponding indeterminacy of spatiotemporal location of any interactions occurring. The singularities predicted by GR lie outside the theory's remit, yet it seems clear that the question of the behaviour of matter-energy fields in the neighbourhoods of such singularities requires quantum treatment. This means that in default of a consistent theory based on the principles of GR and QT, or at least a theory that delivers these principles as results under certain constraints or approximations, we have at best a phenomenological understanding of the big bang and of black holes. These facts alone would motivate the quest for a

---

[39] See Bell ([1981b] 1987, 146), and for thorough discussions of Bell's theorem (Maudlin 2002) and (Shimony 2017).

[40] A similar view is espoused by Abner Shimony in his (2017), where he suggests a return to Heisenberg's idea of probability amplitudes as potentialities. But for the reasons stated, I cannot agree with him that this situation involves "violat[ing] the restrictions that space-time structure imposes upon actual events", or that the correlations can properly be said to exhibit a "peculiar kind of causality".

[41] I am indebted to my colleague Barry Allen for this formulation.

theory of quantum gravity, even though there are other disparate motivations in play, such as the attempt to unify gravity with the other four fundamental forces, or the desire to avoid postulating a 4-D continuum. But for all the fundamentality of such theoretical considerations, the available empirical evidence bearing on them is slight and necessarily indirect: the arena in which quantum corrections of GR would come into play is characterized by the Planck length, $l_P = \sqrt{\hbar \mathrm{G}/\mathrm{c}^3} \approx 1.616 \times 10^{-35}$ m. As Jeremy Butterfield has noted, this is "as many orders of magnitude smaller than the proton, as the proton is smaller than the Earth!" (Butterfield 2002, 325).

There are various different approaches to quantum gravity,[42] but my concern here is with those that take time to be dispensable. This is held to be the inescapable conclusion of approaches which quantize General Relativity in terms of a Hamiltonian formalism with constraints (to be explained below). These approaches are known collectively as Canonical Quantum Gravity (CQG), and result in the infamous Wheeler-DeWitt equation, an equation that appears to characterize a universe in which there is no change in time. Two of the foremost proponents of CQG have been Julian Barbour and Carlo Rovelli, who argue that time is not fundamental, but at best a derived concept, with the flow of time rendered either an illusion (Barbour) or something existing only relative to a human observer's macroscopic perspective (Rovelli).

Clearly, the idea that there is no time in the foundations of nature is contrary to ordinary experience, so much so that if it is not to be regarded as empirically refuted it will require a complete reinterpretation of what it means to experience time or temporal succession. Accordingly, I will analyze these authors' theories in two stages. First I will sketch the main features common to their shared approach, eschewing technical details as far as is possible, and raise objections to some of their founding assumptions. Then I will turn to their respective attempts to reconcile their theories with experience, critiquing these in turn.[43]

Two founding assumptions in CQG are that there is such a thing as the wave function of the universe, and that the spacetime of general relativity must be treated in a "3 +1 decomposition", that is, split into a succession of three-dimensional spacelike hypersurfaces. In quantum theory, the evolution of any system is given by the Schrödinger equation, where the Hamiltonian function generates successive states of the system in time. The "problem of time" is then that in GR there is no preferred time coordinate (or way of *foliating* spacetime). Because Einstein's field equations are generally covariant, there is no preferred time coordinate that could track the evolution of the system in time. CQG theories take what is called

---

[42] See (Kiefer 2011) and (Weinstein and Rickles 2018) for clear surveys of approaches to quantum gravity. A notable lacuna in my critique is the absence of any analysis of one of the chief claims of Loop Quantum Gravity, namely that spacetime itself must be quantized. This is simply too big an issue to tackle in a work of this kind.

[43] My procedure is thus complementary to that of Karim Thébault in his (2012), who puts aside philosophical difficulties of interpretation to concentrate on issues of a more technical, mathematical-physical nature, for which his thorough article should be consulted.

a constrained Hamiltonian approach to quantum gravity.[44] In broad strokes, this can be explained as follows. The Hamiltonian for the whole universe can no longer generate time translations, but must instead generate simply changes of foliation and spatial coordinate transformations. These do not, however, generate any discernible physical consequences, so the Hamiltonian acts only as a constraint on the other dynamical variables. A technique for treating such redundant variables in quantum theory was introduced by Dirac: restrictions are placed on the Hilbert space on which the wavefunction is defined, and this leads also to restrictions on the quantum mechanical operators. These restrictions are dire: they entail that the only natural operators that are definable are constants of the motion. In the words of Bill Unruh, "Only those features of the spacetime geometry which do not change from arbitrary time to arbitrary time seem implementable as operators on the Hilbert space. Thus the dynamical content of the theory seems to be trivial—nothing changes" (Unruh 1995, 57–58).

If this seems rather abstract, Barbour (1999, 241–2) has given a very neat (he calls it "simple-minded") way to make this point within his Machian framework. The classical Schrödinger wave function ψ for a system of N quantons will in general change if any of the following variables is changed: the relative configuration C of the quantons in configuration space, the location of the centre of mass of the system M, the orientation of the system O, or the external time at which we evaluate the function t. We may thus represent the wave function as $\psi(C, M, O, t)$, and the (time-dependent) Schrödinger equation governing its evolution in time will be

$$\hat{\mathrm{H}}\psi(C, M, O, t) = i\hbar\frac{\partial}{\partial t}\psi(C, M, O, t)$$

But now suppose we consider the wave function of the whole universe. On the Machian view Barbour favours (where rotations and positions are always relative), there is no way to change either the position of the centre of mass or the orientation of the universe: since the universe is all there is, with respect to what could its position or orientation be defined? Similarly, if there is nothing outside the universe, then there is no uniform process there by which to define an external time. So the wave function of the universe Ψ can have no rate of change in time, and must depend only on the relative configurations C. It will be governed by the time-independent Schrödinger equation known as the Wheeler-DeWitt equation:

$$\hat{\mathrm{H}}\psi(C) = 0$$

Now, time-independent Schrödinger equations are hardly new. In fact, as Barbour observes (1999, 230), Schrödinger discovered this form of equation first, before the time-dependent one. The time-independent equation is used to find all the stationary states a given system may have if its wavefunction undergoes no change. Each

---

[44] The earliest of these canonical quantum gravity theories is Quantum Geometrodynamics. Barbour's and Bertotti's "best-matching" approach (Barbour and Bertotti 1982) and the Loop Quantum Gravity of Smolin, Rovelli and Ashtekar (1992) are later approaches.

such state corresponds to a definite energy and frequency or wavelength, like the static wave forms of differing wavelengths you can produce by pulsing one end of a skipping rope with the other remaining fixed. A system like Bohr's atom exists in a superposition of such stationary states, creating the characteristic interference terms, and yielding probabilities of a transition from one state to the other with a drop in energy corresponding to the difference in frequencies of the two states in question, with this energy carried off in the emitted photon. In the wave function of the entire universe, however, the whole wave function represents a stationary state, according to the Wheeler-DeWitt equation. But if change in quantum theory is a consequence of interference between states of differing energies, then, if the universe itself is in a stationary state, there can be no change in it (Barbour 1999, 231).

Before proceeding to some of the specifics of Barbour's and Rovelli's arguments, let me note some general objections that apply equally to all canonical theories of quantum gravity. First, they depend explicitly on the idea of the wave function of the universe, the supposition of which was critiqued above. Second, they lay themselves open to all the objections that have been made of the MWI. There are the difficulties with probability: how could such a theory be tested, there being no ensemble of "similarly prepared" universes available to inspection; in the words of Fay Dowker: if "all possibilities are realized, there is no role for probabilities".[45] In fact, even this objection assumes, with the MWI, that the possibilities represented by the correlated states *are* the events that are actualized. But as I objected above, this obscures the very meaning of "event".

Second, the CQG theories are premised on the decomposition of spacetime into three-dimensional spacelike hypersurfaces, as in the FLRW solutions for the Einstein field equations. As we saw in the previous chapter, these solutions are appropriate for the very large-scale structure of the universe, where abstraction has been made away from all local irregularities. But in these solutions there is no such thing as a "state of the universe at a time" except in this large-scale, abstract sense: cosmic time does not track the evolution of any local processes. (Even for an object as "small" as the Earth-Sun system, the Schwartzschild solution must be used in order to represent the local field of the Sun.) Perhaps this can be circumvented by the shape dynamics approach, but one still wonders at the appropriateness of using a decomposition into space and time for a general, quantum theory of gravity.

Both Barbour and Rovelli, however, are sufficiently convinced that the problem with time is unavoidable given the different roles time plays in quantum theory and in general relativity, and therefore inescapable in any formulation of quantum gravity.[46] Let me now turn to their arguments.

---

[45] See Fay Dowker's exchange with Barbour quoted by him, to his credit, in (Barbour 1999, 355). Barbour's reply is that Born's probability density is recorded in his Platonia as an incredibly strong concentration "on a tiny proportion of its points that all turn out to be time capsules as I define them" (356).

[46] Cf. Barbour (1999, 310): "The timelessness of the Wheeler-DeWitt equation, found by well-tried quantization methods, reflects the deepest structure of Einstein's theory [sc. general relativity]." Rovelli endorses the Wheeler-DeWItt equation, holding that "the conventional structure of QM is *certainly* physically incomplete, in the light of GR" (Rovelli 2009, 6).

Barbour's approach is Machian. That is, he follows Mach in eschewing absolute space and time. He remains faithful to a relationalism about space, conceiving it solely in terms of relations of distance among bodies, and time in terms of relations among spatial configurations. The aim is to build the universe out of "places", "where 'place' means a relative arrangement, or configuration, of the complete universe" (Barbour 1999, 69). I will not attempt to duplicate Barbour's subtle arguments here—I refer the reader to his own eminently readable account in (Barbour 1999). But the following rough sketch gives, I hope, the main elements of his construction. The key ideas are *relative configuration*, *best-matching* and *distinguished simplifier*. A "relative configuration" is a representation of the universe at an instant in terms of the relative positions of bodies (or points in a field) to one another, the totality of which Barbour calls "Platonia". Now, for a relative configuration of a number of bodies you can define a certain quantity d in terms of their masses and mutual distances, and then define the *intrinsic difference* between them as the unique quantity obtained by minimizing d. This gives their *best-matching position*. Knowing all these relative distances in configuration space, you can "determine the geodesics in Platonia that correspond to classical Machian histories" (Barbour 1999, 170). In terms of these you can then define a *Machian distinguished simplifier*, which gives a measure of how much change has occurred in passing from one best-matching position to another slightly different one: "the time separation that unfolds the dynamical history in the simplest or most uniform way" (170). This distinguished simplifier can then be used as a surrogate for duration giving a unique value for the distance apart in time of two events occurring at different Nows (Barbour 1999, 120, 170). In classical physics, it is the ephemeris time that performs the role of distinguished simplifier.[47]

Working with Bruno Bertotti, Barbour was able to make a similar construction work for general relativity. Spacetime is modelled as a sequence of Nows, and best-matching is used to define the distance between them in relative configuration space, yielding the geodesics of GR: proper time is recovered as a local ephemeris time. Finally, in Barbour's completion of this theory by incorporating quantum considerations, Platonia becomes the space of configurations of relative states, keyed to decompositions into position eigenstates. Histories of individual processes are unique sequences of states in Platonia. But there is no time outside the universe to time such changes. Rather, "changes are just what take the universe from one place in Platonia to another" (Barbour 1999, 69), and all changes are on an equal footing for the purposes of timing. Change, that is, is simply difference of one configuration from another, and "we reckon time by the totality of changes" (69). It is a mistake to think of things taking different speeds along their paths through Platonia: "With time gone, motion is gone" (69). "The history of the universe *is* the path" (69), any point on the path being simply a configuration of the universe, a "time capsule" (30 ff.).

Barbour introduces his notion of "time capsule" in order to try to account for the appearance of time. On this account histories are all illusions, created by records of the past contained in the present moment. A time capsule is "any fixed pattern that creates or encodes the appearance of motion, change or history" (30), or more

---

[47] See Chap. 4 above for a discussion of ephemeris time.

formally, “any static configuration that appears to contain mutually consistent records of processes that took place in a past in accordance with certain laws” (31). In formulating this account, Barbour avails himself of an interpretation of Everett’s relative state interpretation of quantum theory given by John Bell in his 1981 essay “Quantum mechanics for cosmologists” (Bell [1981a] 1987, 117–138). So let me briefly sketch that first.

There are three main elements to Bell’s reconstruction of Everett’s position. (1) He sees it as having in common with his favoured de Broglie-Bohm theory, not just the rejection of collapse, but also the preferential nature of position measurements. This follows from Everett’s endorsement of Bohm’s analysis of what came to be called decoherence. (2) In keeping with his rejection of the collapse of the wave packet, Everett had contended that, for the standpoint of the theory, “*all* elements of a superposition (all ‘branches’) are ‘actual’, none any more ‘real’ than the rest”, and that “no observer will be aware of any ‘splitting’ process” (Everett 1957, 495, n.‡). Despite this talk of “splitting”, however, Bell contends that the various paths through the configuration space are all “there” prior to such an instant as well as after it. For “at the microscopic level there is no such asymmetry in time as would be indicated by the existence of branching and non-existence of debranching” (Bell [1981a] 1987, 135). And this entails, Bell argues, that the wave function of the universe “does not associate a particular branch at the present time with any particular time in the past, any more than any particular branch in the future” (135). For him this is “the really novel element” in Everett’s theory, “a repudiation of the concept of the ‘past’, which could be considered in the same liberating tradition as Einstein’s repudiation of absolute simultaneity” (118). (3) The third and “essential” element Bell identifies is the claim that “this does not matter at all. For we have no access to the past. We have only our ‘memories’ and ‘records’.” (136). Why “memories” and not just objective records? Because, on Everett’s view, the observer is actually *perceiving* the states on one of the branches. As he wrote in 1957, when an isolated system of two components (such as the cat and the observer) is in a superposition, “the object-system state is a particular eigenstate of the observation”,[48] and “*furthermore the observer-system state describes the observer as definitely perceiving that particular system state*” (Everett 1957, 459; 1983, 320). This is, of course, a very strange use of the word “perceive”; we’ll come back to that.

Barbour takes all three elements of Bell’s reconstruction of Everett’s position—the preferential nature of position eigenstates,[49] the repudiation of the past, and the interpretation of certain configurations as including observers’ experiences—as the inspiration for his attempt to account for the persistent illusion that we live in time. Instead, there are just time capsules, “static configurations”, each one a record encapsulating

---

[48] Correspondingly, Everett claims that the wave function of the combined system is in a superposition of states “for each of which the apparatus has recorded a definite value”, say α_i. As Torretti objects, this is not true: rather, the apparatus “*would record* the definite value α_i *if* the compound system *were in that state*—which it is not: it is in a superposition” (Torretti 1999, 390).

[49] Cf. Barbour (1999, 301): “Bell advocated a simple and robust answer to this [sc. the preferred basis problem]…: the complete system formed by the particles and the instruments measuring them is always defined in the last resort by positions”.

a whole (but illusory) history, including the memories of anyone experiencing it. A time capsule is an experienced instant, spread out through the whole of space:

> Is not our most primitive experience always that we seem to find ourselves, in any instant, surrounded by objects in definite spatial positions? Each experienced instant is thus of the nature of an observation, a discovery, even—we establish *where we are*. Moreover what we observe is always a collection, or totality, of things. We see many things at once. (Barbour 1999, 265)

I think there are some serious objections to Barbour's position, which I hope will be apparent to anyone who has followed the arguments of the preceding chapters.

First, we do not experience all of space in an instant. Any perception is necessarily temporally extended, and we perceive bodies in spatial relationships to us as a result of stable patterns revealed in images built up over time. The present relative to an extended event like an act of perception has a spatiotemporal extent, as argued in Chaps. 3 and 6; what is more, if what we *experience* is a measure, it includes events in the whole backward light-cone. But it is a cardinal error to try to construct reality out of items of experience (Chap. 6). Moreover, the idea of spatial location at an instant is fundamentally non-relativistic: this is why there are no Lorentz-invariant spatial location operators in relativistic quantum theory. As for a time coordinate spanning the whole universe, like the cosmic time in FLRW spacetime, this has no relation to individuals' experiences; nor, if the individual is rotating, is it even possible to conceive spacetime as built up from simultaneity slices in that observer's frame of reference, as Dieks has argued (Chap. 7).

Second, there is Bell's claim that Everett's MWI entails the "repudiation of the concept of the 'past'". This is central to Barbour's conception. If there were time (call this statement p), then (q) trajectories through configuration space would represent possible histories. In the case of Mott's α-particle emission described above, "an immense number of wave-function 'fingers' emerge almost at once and race in a multitude of directions across configuration space" (Barbour 1999, 294). But "at the microscopic level" there is no unique trajectory through a point in configuration space (not-q), Bell claims, since "branches" can reunite, as in cases of interference. So, Barbour can be seen to conclude by Modus Tollens, (not-p), there is no time. Change, as we quoted him saying above, is mere difference. But this is the "static" theory of time, defended by Russell and others, that was subjected to criticism in Chaps. 2 and 3. A "B theory" of time that does not recognize the intrinsic directionality of process becomes a McTaggartian "C theory", where nothing happens at all. It is the becoming of later states out of earlier ones that is responsible for the directionality of process, so to deny passage in this sense, to deny that processes have such directionality—even at the microlevel—is to deny motion. Barbour, unlike many "B theorists", is at least consistent on this point: "With time gone, motion is gone" (Barbour 1999, 69), he says, equating passage with time itself.[50] "When we think we see motion at some instant," he claims, "the underlying reality is that our brain at that instant contains data

---

[50] "B Theorists" like Smart and Grünbaum want to deny passage while still upholding the reality of time and motion. But, as I argued in Chaps. 2 and 3, the reality of passage stands or falls with the reality of motion.

corresponding to several different positions of the object perceived to be in motion" (1999, 266–267). That's all well and good, but to claim that the brain, "through the way in which it presents data to consciousness, somehow 'plays the movie' for me" (267) is to use the language of passage in an attempt to explain away passage, as I argued in Chap. 2.[51]

The third element Barbour appeals to is Everett's strange notion that "the observer" has direct awareness of—"perceives"—certain states in configuration space. "What we take to be wave-function collapse," Barbour declares, "is merely finding that this ineffable self-sentient something that we call ourselves is in one point of the configuration space rather than another" (1999, 297–98). Indeed, "all observation, which is simultaneously the experiencing of one instant of time, is ultimately a (partial) locating of ourselves in Platonia" (298). This is why Barbour approves "the really novel element" that Bell saw in Everett's theory, the replacement of the past by our present memories of the past. He repeats Bell's words: "Our only evidence of the past is through present records. If we have them, the actual existence of the past is immaterial" (Bell [1981a] 1987, 136; Barbour 1999, 300). But there Bell is only reconstructing Everett's reasoning—he himself will have none of it: "Everett's replacement of the past by memories is a radical solipsism," Bell argues, "extending to the temporal dimension the replacement of everything outside my head by my impressions" (136). Barbour demurs. But Bell is surely right. It is a typical idealist fallacy to identify the *existence* of something (an ontological matter) with our *evidence for its existence* (a matter of epistemology).

Rovelli's approach differs from Barbour's in that it is based on *Loop Quantum Gravity*, a theory of which he (along with Lee Smolin and Abhay Ashtekar) is one of the chief proponents—see, e.g. (Ashtekar et al. 1992). What it has in common with Barbour's, though, is its basis in the Hamiltonian-with-constraints approach, which leads to the Wheeler-DeWitt equation, and its apparent implication that there is no dynamical change in the universe. His reaction to this is very different from Barbour's, however. For him, "the fact that we cannot arrange the universe [as] a single orderly sequence of times does not mean that nothing changes" (Rovelli 2018, 109). It is just that "the temporal structure of the world is more complex than a simple single linear succession of instants" (109). Although there is no time governing the whole universe, there are, he asserts, intrinsic times, each corresponding to a dynamical variable. These times are strictly relational: the rate of change of a dynamical variable is simply a relation between its changes and those of another taken as standard. In Rovelli's view mechanics itself—classical and general relativistic, as well as quantum—is "a theory of relations between variables, rather than the theory of the evolution of variables in time" (Rovelli 2008, 1).

Again, I will not attempt to give an account of Rovelli's physical theory (I doubt I am competent to do so in any case), but after the briefest of sketches, I will restrict

---

[51] Ellis raises the same objection: "Barbour claims [7, 12] there exist records of events that our brains read sequentially, and so create a false illusion of the passage of time. Thus brain processes are responsible for illusion of change. But 'processes' are things that unfold in time!—there are no processes unless time flows. You can't perceive a flow of time unless time flows, because perception is a process that takes place in time." (Ellis 2012, 20).

my critical remarks to his attempt to reconcile that theory with our experience of time and its passing.

Perhaps the main feature of Loop Quantum Gravity that distinguishes it from rival approaches like String Theory, is that is rejects the idea of a background spacetime. It takes the chief lessons of general relativity to be that location is relational and that spacetime itself is dynamical, and the chief lesson of quantum theory to be that canonically conjugate dynamical variables are subject to Heisenberg's indeterminacy. I will not attempt to describe exactly what loops are, mathematically. Suffice to say that if they are smooth and non self-intersecting, they can be used to solve the Wheeler-DeWitt equation. As described by Rovelli, the theory builds up space from quanta of space into what are called "spin networks", a ring in such a network being the "loop" that gives the theory its name (Rovelli 2018, 126). These networks are webs of interactions, interactions that presuppose only an "elementary form of time that is neither directional nor linear" (Rovelli 2018, 124–125). All these interactions, moreover, are probabilistic and relational: "everything that happens in the world", all events—"including among them the passage of time—are always triggered by an interaction with, and with respect to, a system involved in the interaction" (125).

This talk of *interactions*, however, must be treated with some caution. That they are relational and probabilistic is a sign of the close affinity of Rovelli's relational quantum theory with Everett's relative state formulation. As for Everett, so for Rovelli, a quantum event is not actually an event, but an eigenstate.[52] There is no suggestion that Rovelli will go along with Everett and Barbour's idea that we perceive eigenstates, or that the observer's experiences are somehow encoded in them. All the same, with respect to becoming, his approach seems to have the same shortcoming as the MWI: the eigenstates that are part of a given decomposition only represent potentials of things to happen or variables to take on certain values: they are not events. This is, of course, consistent with the claim that there is no time in the foundations of things, but not with Rovelli's claims that the world is not composed of things but of events, happenings, and that "the destruction of the notion of time in fundamental physics is the crumbing of the first of these two perspectives, not the second" (Rovelli 2018, 97).

That consideration is relevant also to Rovelli's attempts to recover features of time as received in experience. Among these are the flow of time, remembering the past and not the future, and the irreversibility of process-tokens in time. Of these Rovelli says "I think that these features are not mechanical. Rather they emerge at the thermodynamical level" (Rovelli 2008, 7). This is Rovelli's "thermal time hypothesis", according to which "what we call 'time' is the thermal time of the statistical state in which the world happens to be, when described in terms of the macroscopic parameters we have chosen." (Rovelli 2008, 3). It is relative to an "interaction" with us on a macroscopic scale that the directionality and flux of temporal processes can be recovered. They are not illusory or subjective, but perspectival: "a generic macro-

---

[52] "Common eigenstates $|s>$ of a complete set of commuting partial observables are denoted quantum events", a "partial observable" being what is represented by a self-adjoint operator in a 'rigged Hilbert space' (Rovelli 2009, 5).

scopic state determines a time” (Rovelli 2018, 136). There is not the space here to give a detailed analysis of Rovelli’s hypothesis. But it would seem vulnerable to the same objections I raised in Chap. 4 against Boltzmann’s attempts to derive the direction of time from statistical considerations: The direction of time cannot be defined in terms of increasing entropy without presupposing temporally directed processes, nor, if microprocesses are held to be symmetric in time, could they compose to form macroprocesses.[53] If I am right, what Rovelli terms “these very peculiar features” of time are rooted in the succession of events as they occur, they are not features of “the time variable”, as he regards them. That process-tokens are future-oriented locally is not a consequence of thermodynamics, but is presupposed by it.

Rovelli gives independent arguments for his claim that the flow of time is not to be found on the mechanical level. He claims that in (non-relativistic) classical mechanics it is also the case that “what we measure is only relative evolution between variables”; it is simply a matter of *convenience* to assume an absolute time with respect to which all quantities evolve, so as to make the equations simpler.[54] But as we saw in Chap. 3, that is more nearly Barrow’s position. Newton’s absolute time is a unique “distinguished simplifier”, to use Barbour’s term. But there is nothing conventional about it if, as Barbour remarks, we wish to be able to “keep our appointments” (Barbour 1999, 170). That a given dynamical process can be used as a surrogate for time (a “relative time”, in Newton’s terms) does not make time dispensable.[55] Similarly, as we saw in our discussion of Special Relativity in Chap. 6, while it is true that time dilation is a relative effect for any two frames in relative uniform motion, this does not make it a purely frame-relative effect in general. There is a real difference in the proper times elapsed for the twins in the twin paradox scenario: one is objectively older than the other, and time has flowed at different rates along these different curves in spacetime joining the events of their departure and reunion.[56]

Similarly in General Relativity the lack of a distinguished time coordinate (or foliation of spacetime) does not entail a universal relativity of motion, as we saw in Chap. 7. Transformations of coordinates do not change the physical facts about what are the sources of the gravitational field or about which bodies are performing inertial motions. The rate of flow of a process is directly affected by gravity, as we

---

[53] Rovelli writes, for example: “If I observe the microscopic state of things, then the difference between past and future vanishes” (2018, 32–33). Ellis also has some strong arguments against “the idea that time-reversible Hamiltonian dynamics provides the foundation for physical theory in general and gravitation in particular” (1), emphasizing that in realistic settings physical processes are no longer reversible due to dissipation of heat energy into thermal fluctuations (Ellis 2012, 6–7).

[54] “But it turns out to be convenient to *assume*, with Newton, that there exist a background variable t, such that all observables quantities evolve with respect to it, and equations are simple when written with respect to it” (Rovelli 2008, 3).

[55] George Ellis makes the same point. That we can give a time-independent formulation “does not mean that time does not flow, it just means that the results of times flow are correlations between relevant variables” (2012, 11). The same applies to the time-independent formulations in quantum theory that Barbour appeals to in arguing for the elimination of time.

[56] Cf. Ellis (2012, 12): “The relative flow of time along different world lines may be different: that is the phenomenon of time dilation, caused by the varying gravitational potentials represented by the metric tensor [78]. But this does not mean it is not well defined along each world line.”

saw. But while it is true that "Every phenomenon that occurs has its proper time, its own rhythm" (Rovelli 2018, 16), and that this rhythm will be different from the perspective of other processes with their own rhythms, this does not entail that such "rhythms" are merely relative. The natural rhythm of a process corresponds to its mass-energy, so that the natural frequency associated with an electron is $\nu_E = m_E c^2/h$, where m_E is the rest mass of the electron. The rest mass is an invariant: it does not vary depending on the frame from which it is viewed, although its effective mass will. Similarly, although from the perspective of a frame moving at a speed v with respect to it, that frequency will manifest as $\gamma\nu_E = \gamma m_E c^2/h$—a higher frequency since $\gamma > 1$—this does not change the rest mass or the natural frequency of the electron. (The fact that Weyl's unified theory of gravity and electromagnetism did not respect this is why it was rejected, as we saw in Chap. 7). An electron (or anything else) moving along a geodesic is moving inertially, and this inertial motion encodes a natural measure of time flow that must be respected in any theory of quantum gravity.

8.6 Summary

- In this chapter we have examined the relevance of quantum theory to the reality of becoming. Section 8.1 gave an overview of quantum theory and its origins in puzzles concerning the interaction of matter and radiation, showing how it has been bedevilled by problems of interpretation since its inception. In Sect. 8.2, we examined the construal of becoming by Reichenbach and others in terms of a transition from probabilities of events to events that actually occur. This led to a discussion of the "measurement problem" and the "collapse of the wave function". It was argued that it is one thing to reject von Neumann's projection postulate, and another to interpret all probable states as actualities, as does the Many Worlds Interpretation. This has the effect of undermining the objective probabilities of the theory, and of leading to an indefensible notion of "event".
- A second "no-collapse" interpretation is the subject of Sect. 8.3, the de Broglie-Bohm theory. This has been championed in particular by John Bell, for the light it throws on quantum entanglement. This continued entanglement of states of systems that have interacted is a chief characteristic of quantum theory. On the basis of the theorem bearing his name, Bell has claimed that it entails the existence of non-local influences between the systems: statistical correlations occur across space-like distances despite having no common cause. Against this, it was argued that such influences cannot be interpreted as processes, since they are correlations among relative states, not events. This does not explain how such nonlocal co-ordinations among states arise, but it does show that they are not in violation of SR.
- Finally in Sect. 8.4 we examined the arguments of Barbour and Rovelli that time is eliminated in (canonical) theories of quantum gravity. Their approach is criticized for its dependence on the idea of a state function for the whole universe, and on treating spacetime in terms of spatial configurations-at-an-instant. Sketches are

then given of their differing attempts to explain the alleged elimination of time in canonical quantum gravity, and of each author's attempt to explain the recovery of the manifest features of time, such as its flow and directionality, and subjected to criticism based on the arguments of previous chapters.

References

Ashtekar, Abhay, Carlo Rovelli, and Lee Smolin. 1992. Weaving a Classical Geometry with Quantum Threads. *Physical Review Letters* 69: 237–240.

Barbour, Julian B. 2008. *The Nature of Time*. http://www.platonia.com/nature_of_time_essay.pdf.

Barbour, Julian B. 1999. *The End of Time: The Next Revolution in Physics*. (2001 reprinting). Oxford: Oxford University Press.

Barbour, Julian B. and Bruno Bertotti. 1982. Mach's Principle and the Structure of Dynamical Theories. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 382 (1783), 295–306.

Bell, J.S. [1964] 1987. On the Einstein–Podolsky–Rosen Paradox. *Physics* 1: 195–200; reprinted in Bell 1987, 14–21.

Bell, J.S. [1981a] 1987. Quantum Mechanics for Cosmologists, pp. 611–37. In *Quantum Gravity*, ed. C. Isham, R. Penrose and D. Sciama. Oxford: Clarendon Press, 1981; reprinted in Bell 1987, 117–138.

Bell, J.S. [1981b] 1987. Bertlmann's Socks and the Nature of Reality. *Journal de physique*, Colloque C2, suppl. au numero 3, Tome 42, 41–61; as reprinted in Bell 1987, 139–158.

Bell, J.S. 1987. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Bohm, David. 1951. *Quantum Theory.* Englewood Cliffs: Prentice Hall; pagination keyed to the reprinting of Chapter 22, §§15–19, in (Wheeler and Zurek (eds.) 1983, 356–368).

Bohm, David. 1952. A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden Variables', I and II. *Physical Review* 85: 166–193; pagination keyed to its reprinting in (Wheeler and Zurek (eds.) 1983, 369–402).

Bohm, David. 1957. *Causality and Chance in Modern Physics*. Philadelphia: University of Pennsylvania Press.

Bohr, Niels. [1928] 1983. The Quantum Postulate and the Recent Development of Atomic Theory *Nature* 121, 580–90 (1928); pagination keyed to its reprinting in Wheeler and Zurek (eds.) 1983, 87–126.

Bohr, Niels. 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review* 48, 692–702 (1935); pagination keyed to its reprinting in Wheeler and Zurek (eds.) 1983, 145–151.

Bohr, Niels. 1958. *Atomic Physics and Human Knowledge*. New York: Wiley.

Borges, Jorge Luis. 1964. *Labyrinths*, ed. Donald A. Yates & James E. Irby. New Directions Publishing.

Born, Max. 1949. *Natural Philosophy of Cause and Chance*. Oxford: Clarendon Press.

Born, Max. 1968. *My Life and My Views*. New York: Charles Scribner's Sons.

Bub, Jeffrey. 1997. *Interpreting the Quantum World.* Cambridge: Cambridge University Press.

Butterfield, Jeremy. 2002. The End of Time? *British Journal for the Philosophy of Science* 53 (2): 289–330.

De Broglie, Louis. 1927. La structure de la matière et du rayonnement et la mécanique ondulatoire. *Comptes Rendus* 184: 273–274.

Deutsch, David and Lockwood, Michael. 1994. The Quantum Physics of Time Travel. *Scientific American*, March 1994, 68–74.

DeWitt, Bryce. 1970. Quantum Mechanics and reality. *Physics Today* 23 (9): 30–35; pp. 155–165 in DeWitt and Graham, eds., 1973.

DeWitt, Bryce, and Neill Graham (eds.). 1973. *The Many Worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.

Dorato, Mauro, and Michael Esfeld. 2010. GRW as an Ontology of Dispositions. *Studies in History and Philosophy of Modern Physics* 41: 41–49.

Einstein, Albert, Boris Podolsky and Nathan Rosen. 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? Originally published in the *Physical Review* 47: 777–780 (1935); 138–143 in Wheeler and Zurek (eds.) 1983, to which it is paginated.

Ellis, George F.R. 2012. Spacetime and the Passage of Time. ArXiv:1208.2611v4 [gr-qc] 26 Aug 2012.

Ellis, George F.R. 2014. The Evolving Block Universe and the Meshing Together of Times. *Annals of the New York Academy of Sciences* 1326 (1), https://arxiv.org/pdf/1407.7243.

Everett, Hugh, III. 1957. 'Relative State' formulation of quantum mechanics. *Reviews of Modern Physics* 29: 454–462; reprinted in Wheeler and Zurek (eds.) 1983, pp. 315–323.

Ghirardi, Giancarlo. 2018. Collapse Theories. *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), ed. Edward N. Zalta, https://plato.stanford.edu/archives/fall2018/entries/qm-collapse.

Heisenberg, Werner. [1927] 1983. Über den anschaulichen Inhalt der quantentheoretishcen Kinematik und Mechanik. *Zeitschrift für Physik* 43: 172–198; trans. J.A. Wheeler and W. H. Zurek as "The Physical Content of Quantum Kinematics and Mechanics", pp. 62–84 in (Wheeler and Zurek eds., 1983).

Hilgevoord, Jan. 1996. The Uncertainty Principle for Energy and Time. *American Journal of Physics* 64 (12): 1451–1456.

Hilgevoord, Jan and Atkinson, David. 2011. Time in Quantum Mechanics. In *The Oxford Handbook of Philosophy of Time*, ed. Craig Callender, 478–490. Oxford: Oxford University Press.

Hoefer, Carl. 2016. Causal Determinism. In *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), ed. Edward N. Zalta. https://plato.stanford.edu/archives/spr2016/entries/determinism-causal/.

Jammer, Max. 1974. *The Philosophy of Quantum Mechanics*. New York: Wiley.

Kiefer, Claus. 2011. Time in quantum gravity. In *The Oxford Handbook of Philosophy of Time*, ed. Craig Callender, 491–503. Oxford: Oxford University Press.

London, Fritz and Edmond Bauer. 1983. The Theory of Observation in Quantum Mechanics, trans. of *La théorie de l'observation en mécanique quantique* (Paris: Hermann, 1939), by A. Shimony, J. A. Wheeler and W. H. Zurek. pp. 217–259 in (Wheeler and Zurek (eds.) 1983).

Maudlin, Tim. 2002. *Quantum Non-Locality and Relativity*, 2nd ed. Oxford: Blackwell Publishing.

Mott, Nevill. 1929. The Wave-Mechanics of Alpha-Ray Tracks. *Proceedings of the Royal Society of London* A126: 79–84; as reprinted in Wheeler and Zurek (eds.) 1983, pp. 129–134.

Peacock, Kent A. 2018. *Quantum Heresies*. Milton Keynes, UK: College Publications.

Penrose, Roger. 2016. *Fashion, Faith and Fantasy in the New Physics of the Universe*. Princeton and Oxford: Princeton University Press.

Przibram, K. (ed.). 1986. *Letters on Wave Mechanics*. New York: Philosophical Library.

Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley: University of California Press.

Robertson, H.P. 1929. The Uncertainty Principle. *Physical Review* 34: 163–164; pagination keyed to its reprinting in (Wheeler and Zurek (eds.) 1983, 127–128).

Rovelli, Carlo. 2009. Forget time. arXiv:0903.3832v3[gr-qc]27 Mar 2009.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Schrödinger, Erwin. 1935a. Discussion of Probability Relations Between Separated Systems. *Mathematical Proceedings of the Cambridge Philosophical Society* 31 (4): 555–563.

Schrödinger, Erwin. 1935b. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* 23: 807–812, 823–28, 844–849; trans. John D. Trimmer in ed. Wheeler and Zurek 1983, pp. 152–167, to which it is paginated.

Shimony, Abner. 2017. Bell's Theorem. *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), ed. Edward N. Zalta. https://plato.stanford.edu/archives/fall2017/entries/bell-theorem/.

Smolin, Lee. 2011. A Real Ensemble Interpretation of Quantum Mechanics. asXiv:1104.2822.

Thébault, Karim P.Y. 2012. Three Denials of Time in the Interpretation of Canonical Gravity. *Studies in History and Philosophy of Modern Physics* 43: 277–294.

Torretti, Roberto. 1999. *The Philosophy of Physics*. Cambridge: Cambridge University Press.

Unger, Roberto Mangabeira, and Lee Smolin. 2015. *The Singular Universe and the Reality of Time*. Cambridge: Cambridge University Press.

Unruh, William G. 1995. Time, Gravity and Quantum Mechanics, 23–65 in (Savitt (ed.) 1995). Savitt, Steven F. (ed.) 1995. *Time's Arrows Today: Recent philosophical work on the direction of time*. Cambridge: Cambridge University Press.

Valentini, Antony. 2002. Signal-Locality in Hidden-Variables Theories. *Physics Letters A* 297: 273–278.

von Neumann, John. 1932. *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer, trans. English by Robert T. Beyer as *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1955). Chapters 5 and 6 reproduced in (Wheeler and Zurek, eds., 1983), pp. 549–647, to which it is paginated.

Wallace, David. 2012. *The Emergent Multiverse, Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press.

Weinstein, Steven and Rickles, Dean. 2018. Quantum Gravity. In *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), ed. Edward N. Zalta. https://plato.stanford.edu/archives/win2018/entries/quantum-gravity/.

Wheeler, John Archibald. 1983. Law Without Law (a compilation form pieces originally published in 1979, 1980 and 1981), pp. 182–213 in (Wheeler and Zurek, eds., 1983).

Wigner, Eugene. Remarks on the Mind-Body Question (originally published 1961), pp. 168–181 in (Wheeler and Zurek, eds. 1983).

Zurek, Wojciech H. 2002. Decoherence and the Transition from Quantum to Classical—Revisited. *Los Alamos Science* 27: 1–24.

Chapter 9
Conclusion

Where, then, does all this leave us?

On our familiar view of things, the universe is divided into past, present and future. Events that are past leave their traces, and by so doing constitute part of present reality. Those in the future do not yet exist, and the present, conceived as a global now spread across the whole universe, divides what has been from what is to come. In our experience this changes as more and more events come into being in the now, and then recede into the past.

We have seen how various features of modern physics present severe challenges for maintaining such a view. Because of the relativity of simultaneity, there is no unique now, no unique set of events that are all present to us at a given instant. We have seen how this has prompted many physicists and philosophers to take up a version of the "block universe" view, according to which all events in spacetime simply *are*, without needing to come into being, and to regard the 'now' as merely subjective. To them that conclusion seems only strengthened by considerations from general relativity. Here spacetime is no longer a mere background, but a dynamical structure determined by its contents. Time is only an aspect of spacetime, but no time coordinate has any special privilege. This consideration, along with the assumption of a wave-function of the entire universe and other matters of a more technical nature, has prompted claims that when general relativity and quantum theory are combined into a theory of Quantum Gravity, time is absent altogether.

Such claims have not gone unopposed. The tendency, however, has been to present the issue as one of the reality of *time*, a single all-embracing entity whose existence must be either accepted or denied. If time is real, so it is contended, there must be a way of defending the conception of a universal now, an edge of spacetime beyond which there lies an open or empty future, into which the universe expands by the creation of new events. I have tried to show that the temporal features of the world are much more ramified than this, that time is is not to be identified with some one privileged time coordinate, and that the classical idea of passage as requiring a universal time should be discarded. In particular I have argued that becoming is essentially a local phenomenon tracked by the path-dependent variable proper time, and not tied to the conception of the advance of a universal now. There can

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1_9

be becoming in every region of spacetime whether or not a global time concept is definable, since processes themselves are (local) instances of becoming. Moreover, as Einstein recognized, there are differential rates of becoming for processes taking different paths through spacetime, those rates being faster the further a process is from being inertial.

The attraction of the universal now is nevertheless extremely strong. There are several reasons for this. It is thought that the experience of events in the now is the bedrock of our experience; that without passage of a universal now all events are equally real, and the now becomes merely subjective; and that without such a now, we can make no sense of the openness of the future, or of the possibility that the laws of nature themselves might emerge or evolve. All these motivations are apparent in the recent book by Unger and Smolin, for example, who present just the kind of dichotomy I am resisting. Either we accept "the reality of time" in the form of a progression of present moments of a global time, or we are stuck with the static block universe picture, where "the experience of moments of time and their passage or flow are illusions" (Unger and Smolin 2015, 361).[1]

As we have seen, though, this idea that we "experience" global nows and their passing is simply unfounded. No event that is now according to a global time function—and this applies especially to events in a cosmological context—is a possible item of our experience. But passage does not require that it should be. Passage, I have argued, is local: events come to be out of other events in their neighbourhoods, and processes are constituted by such successions of events or states of a system undergoing evolution. This does not require a global now. Nor does it require that events should be experienced. Our experience of certain events as present can nevertheless be accounted for by a conception of perceptions as temporally extended processes, I have argued. Those extended events and processes contained within a spacetime region centred on the event of our perceiving will then be those with which interaction can occur during the time of perception, with such physical interconnection giving us a robust sense of their reality. Such a present relative to a given extended event (such as our forming a perception) is an extended region of spacetime, the Alexandrov present. At any point within such a present, however, only those events will have become which are in the absolute past (backward lightcone) of that point. As Robb was the first to argue, pairs of events that have no possibility of interacting—each one being in the other's "elsewhere"—have a spacelike relation but no order of succession to one another. In the Minkowski spacetime of special relativity, events that do succeed one another do so in a partial order, Robb's conical order: two spacelike separated events, a and b, are neither before nor after one another, but they will both be in the future of some event in the past of both of them, and in the past of some possible event in their intersecting absolute futures. The same goes for any spacetime obeying the chronological principle.

---

[1] Thus Smolin writes: "Unless we want to retreat to a kind of event or observer solipsism in which what is real is relative to observers or events, we need a real and global notion of the present" (Unger and Smolin 2015, 418). Both authors subscribe to a presentist conception according to which "all that is real is real in a present moment which is one of a series of moments" (Unger and Smolin 2015, 361 and 415).

Another motivation for the global now is to make sense of the openness of the future. Without an objective global now dividing events in the past from an undetermined future, it is alleged, we again fall into the rigid determinism of the static block.[2] But an objective distinction between past and future at any spacetime location does not require a global now splitting all the events in the universe into three classes, those that are past, those that are happening now, and those that are yet to come in an "empty future". The problem with the idea of an open future is the supposition that "the future" has an absolute connotation. It ignores the relational character of the terms "past" and "future".[3] Events that are in my absolute past have happened as of this spacetime location, whereas there is at this location no fact of the matter about anything in its absolute future. But that is the same for any other location! The events themselves happen at the very spacetime location in which they happen. Whether an event is past or future is not an absolute quality of an event, but depends on the vantage point from which it is being considered.

There is, I have argued, no vantage point from which spacetime itself can be considered as coming into being. To regard worldlines as growing by the steady accretion of events is a mistake: a worldline is a 4-dimensional spatiotemporal entity whose representation already includes time, not a 3-dimensional spatial object that can change in time. Spacetime as a whole is not something that can either change or remain the same, since that would require a time outside of spacetime according to which changing or staying the same could be represented. Nevertheless, there is change everywhere in it. Every event, no matter how short its duration, is a process, an instance of becoming. There is therefore no question about whether an event has become: representing it as an event is representing it as having become, since an event *is* something that has become.

So spacetime as a whole does not evolve. The expansion of the universe is a *spatial* expansion. It denotes the fact that world lines representing the mean motion of clusters of galaxies are all moving apart, as evidenced by the red shift in radiation. This is represented in the FLRW solutions to Einstein's field equations, which allow the decomposition of spacetime into space and a cosmic time, the time by means of which we gauge the age of the universe. Now these solutions depend on the viability of Weyl's postulate (that there is a mean motion on an intergalactic scale). I have suggested that this would no longer hold in the very early universe, putting in doubt the very notion of a cosmic time in early stages of the universe. This is already the case at scales far larger than the Planck scale, where quantum effects should start to dominate.

---

[2] According to Smolin, "we need an objective distinction among past, present, and future to be able to assert that there are no certain facts of the matter about the future" (Unger and Smolin 2015, 522).

[3] Oliver Pooley advocates a similar perspectival view in (Pooley 2013), arguing that "it is simply not plausible to take as absolute those facts that correspond to the perspective of a space-time region that is both spatially as well as temporally local" (357). As a criticism of Smolin, the accusation of absolutism is of course highly ironic, given Smolin's oft-stated commitment to a through-going relationalism (see, e.g., Unger and Smolin 2015, 355–356).

In the quantum domain, some physicists have speculated, we should not expect time to be anything like the time of our macroscopic experience. Relativity already shows that time does not pass at a unique rate, but runs faster at the top of Bacon's mountain than at the bottom of his mine, eradicating the notion of a unique rate of flow. Likewise, it is suggested, our intuitive impression that time passes asymmetrically from past to future may not apply at all in the quantum domain. That could be an illusion created by the way our experience is constituted by such micro-processes. In fact, at a really fundamental level, there may be no time at all.[4]

Let me recount why I cannot agree with those contentions. I have argued that Einstein's General Relativity does not license a universal relativity of motion, and that the absence of a unique time coordinate is no argument against there being a standard against which rates of flow may be gauged. The fact that a given dynamical process may stand in for time (thus acting as what Newton called a "relative time"), supports neither the claim that time is eliminable in dynamics, nor the claim that choosing the rate of flow of a process is a matter of convenience. I argued that a similar view implicit in Barrow's work was refuted by Newton, who followed Huygens' lead in taking the time beaten out by an inertial motion as giving a standard for equable flow. The same standard is incorporated into relativity theory as the Geodesic Principle, where a geodesic represents an inertial path, the path of minimum rate of flow.

I have also argued that process is a succession of states or concrete events whose order is intrinsically directed from initial to final state, and that this is what constitutes the direction of time. This makes unintelligible the very idea of a *process* that is not directed in time, let alone the idea that elementary (microscopic) processes with no time orientation could compose to form a time-directed macroscopic process. The irreversibility of the huge majority of processes occurring in the world is another matter, due to the relative improbability of the initial conditions used in applying the equations of physics, but it is not to be sought in the equations themselves.

Finally, there is the problem of accounting for the appearance of events succeeding one another without presupposing a real temporal succession of those appearances. This difficulty, I contend, has not been solved by anyone who denies the passing of time. Time really does flow, in the sense that events and states objectively come after those preceding them in their neighbourhoods, and at rates which vary according to the paths that processes take through the fields defining spacetime.

---

[4] As we have seen, all these contentions are made by Carlo Rovelli in his (2018). Although he grants that a unique quantity "time" in the equations of fundamental physics "does not imply a world that is frozen and immobile" (96), and talks freely of time passing locally at different speeds (194), he insists that rates of flow are purely relative (15–17, 117–120), that time's passing asymmetrically from past to future is only trick of (human) perspective (194), that time is dispensable in the description of the world (117–18), and that "in the elementary grammar of the world there is neither space nor time" (195).

References

Pooley, Oliver. 2013. Relativity, the Open Future, and the Passage of Time. In *Proceedings of the Aristotelian Society*, vol. cxiii, Part 3.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Unger, Roberto Mangabeira, and Lee Smolin. 2015. *The Singular Universe and the Reality of Time*. Cambridge: Cambridge University Press.

Author Index

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1

Subject Index

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1

Titles in The Frontiers Collection

Quantum Mechanics and Gravity
By Mendel Sachs

Quantum-Classical Correspondence
Dynamical Quantization and the Classical Limit
By A. O. Bolivar

Knowledge and the World: Challenges Beyond the Science Wars
Ed. by M. Carrier, J. Roggenhofer, G. Küppers and P. Blanchard

Quantum-Classical Analogies
By Daniela Dragoman and Mircea Dragoman

Quo Vadis Quantum Mechanics?
Ed. by Avshalom C. Elitzur, Shahar Dolev and Nancy Kolenda

Information and Its Role in Nature
By Juan G. Roederer

Extreme Events in Nature and Society
Ed. by Sergio Albeverio, Volker Jentsch and Holger Kantz

The Thermodynamic Machinery of Life
By Michal Kurzynski

Weak Links
The Universal Key to the Stability of Networks and Complex Systems
By Csermely Peter

R. T. W. Arthur, *The Reality of Time Flow*, The Frontiers Collection,
https://doi.org/10.1007/978-3-030-15948-1

The Emerging Physics of Consciousness
Ed. by Jack A. Tuszynski

Quantum Mechanics at the Crossroads
New Perspectives from History, Philosophy and Physics
Ed. by James Evans and Alan S. Thorndike

Mind, Matter and the Implicate Order
By Paavo T. I. Pylkkanen

Particle Metaphysics
A Critical Account of Subatomic Reality
By Brigitte Falkenburg

The Physical Basis of the Direction of Time
By H. Dieter Zeh

Asymmetry: The Foundation of Information
By Scott J. Muller

Decoherence
and the Quantum-To-Classical Transition
By Maximilian A. Schlosshauer

The Nonlinear Universe
Chaos, Emergence, Life
By Alwyn C. Scott

Quantum Superposition
Counterintuitive Consequences of Coherence, Entanglement, and Interference
By Mark P. Silverman

Symmetry Rules
How Science and Nature are Founded on Symmetry
By Joseph Rosen

Mind, Matter and Quantum Mechanics
By Henry P. Stapp

Entanglement, Information, and the Interpretation of Quantum Mechanics
By Gregg Jaeger

Relativity and the Nature of Spacetime
By Vesselin Petkov

The Biological Evolution of Religious Mind and Behavior
Ed. by Eckart Voland and Wulf Schiefenhövel

Homo Novus-A Human without Illusions
Ed. by Ulrich J. Frey, Charlotte Störmer and Kai P. Willführ

Brain-Computer Interfaces
Revolutionizing Human-Computer Interaction
Ed. by Bernhard Graimann, Brendan Allison and Gert Pfurtscheller

Extreme States of Matter
On Earth and in the Cosmos
By Vladimir E. Fortov

Searching for Extraterrestrial Intelligence
SETI Past, Present, and Future
Ed. by H. Paul Shuch

Essential Building Blocks of Human Nature
Ed. by Ulrich J. Frey, Charlotte Störmer and Kai P. Willführ

Mindful Universe
Quantum Mechanics and the Participating Observer
By Henry P. Stapp

Principles of Evolution
From the Planck Epoch to Complex Multicellular Life
Ed. by Hildegard Meyer-Ortmanns and Stefan Thurner

The Second Law of Economics
Energy, Entropy, and the Origins of Wealth
By Reiner Kümmel

States of Consciousness
Experimental Insights into Meditation, Waking, Sleep and Dreams
Ed. by Dean Cvetkovic and Irena Cosic

Elegance and Enigma
The Quantum Interviews
Ed. by Maximilian Schlosshauer

Humans on Earth
From Origins to Possible Futures
By Filipe Duarte Santos

Evolution 2.0
Implications of Darwinism in Philosophy and the Social and Natural Sciences
Ed. by Martin Brinkworth and Friedel Weinert

Probability in Physics
Ed. by Yemima Ben-Menahem and Meir Hemmo

Chips 2020
A Guide to the Future of Nanoelectronics
Ed. by Bernd Hoefflinger

From the Web to the Grid and Beyond
Computing Paradigms Driven by High-Energy Physics
Ed. by René Brun, Frederico Carminati and Giuliana Galli-Carminati

The Language Phenomenon
Human Communication from Milliseconds to Millennia
Ed. by P.-M. Binder and K. Smith

The Dual Nature of Life
Interplay of the Individual and the Genome
By Gennadiy Zhegunov

Natural Fabrications
Science, Emergence and Consciousness
By William Seager

Ultimate Horizons
Probing the Limits of the Universe
By Helmut Satz

Physics, Nature and Society
A Guide to Order and Complexity in Our World
By Joaquín Marro

Extraterrestrial Altruism
Evolution and Ethics in the Cosmos
Ed. by Douglas A. Vakoch

The Beginning and the End
The Meaning of Life in a Cosmological Perspective
By Clément Vidal

A Brief History of String Theory
From Dual Models to M-Theory
By Dean Rickles

Singularity Hypotheses
A Scientific and Philosophical Assessment
Ed. by Amnon H. Eden, James H. Moor, Johnny H. Søraker and Eric Steinhart

Why More Is Different
Philosophical Issues in Condensed Matter Physics and Complex Systems
Ed. by Brigitte Falkenburg and Margaret Morrison

Questioning the Foundations of Physics
Which of Our Fundamental Assumptions Are Wrong?
Ed. by Anthony Aguirre, Brendan Foster and Zeeya Merali

It From Bit or Bit From It?
On Physics and Information
Ed. by Anthony Aguirre, Brendan Foster and Zeeya Merali

How Should Humanity Steer the Future?
Ed. by Anthony Aguirre, Brendan Foster and Zeeya Merali

Trick or Truth?
The Mysterious Connection Between Physics and Mathematics
Ed. by Anthony Aguirre, Brendan Foster and Zeeya Merali

The Challenge of Chance
A Multidisciplinary Approach from Science and the Humanities
Ed. by Klaas Landsman, Ellen van Wolde

Quantum [Un]Speakables II
Half a Century of Bell’s Theorem
Ed. by Reinhold Bertlmann, Anton Zeilinger

Energy, Complexity and Wealth Maximization
Ed. by Robert Ayres

Ancestors, Territoriality and Gods
A Natural History of Religion
By Ina Wunn, Davina Grojnowski

Space, Time and the Limits of Human Understanding
Ed. by Shyam Wuppuluri, Giancarlo Ghirardi

Information and Interaction
Eddington, Wheeler, and the Limits of Knowledge
Ed. by Ian T. Durham, Dean Rickles

The Technological Singularity
Managing the Journey
Ed. by V. Callaghan, J. Miller, R. Yampolskiy, S. Armstrong

How Can Physics Underlie the Mind?
Top-Down Causation in the Human Context
By George Ellis

The Unknown as an Engine for Science
An Essay on the Definite and the Indefinite
By Hans J. Pirner

CHIPS 2020 Vol. 2
New Vistas in Nanoelectronics
Ed. by Bernd Hoefflinger

Life – As a Matter of Fat
Lipids in a Membrane Biophysics Perspective
By Ole G. Mouritsen, Luis A. Bagatolli

The Quantum World
Philosophical Debates on Quantum Physics
Ed. by Bernard D’Espagnat, Hervé Zwirn

The Seneca Effect
Why Growth is Slow but Collapse is Rapid
By Ugo Bardi

Chemical Complexity
Self-Organization Processes in Molecular Systems
By Alexander S. Mikhailov, Gerhard Ertl

The Essential Tension
Competition, Cooperation and Multilevel Selection in Evolution
By Sonya Bahar

The Computability of the World
How Far Can Science Take Us?
By Bernd-Olaf Küppers

The Map and the Territory
Exploring the Foundations of Science, Thought and Reality
By Shyam Wuppuluri, Francisco A. Doria

Wandering Towards a Goal
How Can Mindless Mathematical Laws Give Rise to Aims and Intention?
Ed. by Anthony Aguirre, Brendan Foster and Zeeya Merali

Computer Simulations in Science and Engineering
Concepts – Practices – Perspectives
By Juan M. Durán

Stepping Stones to Synthetic Biology
By Sergio Carrà

Information — Consciousness
Reality How a New Understanding of the Universe Can Help Answer Age-Old Questions of Existence
By James B. Glattfelder

Spirals and Vortices:
In Culture, Nature, and Science
Ed. by KinkoTsuji, Stefan Müller

What is Fundamental?
Ed. by Anthony Aguirre, Brendan Foster, Zeeya Merali

Particles, Fields and Forces:
A Conceptual Guide to Quantum Field Theory and the Standard Model
By Wouter Schmitz

The Reality of Time Flow
Local Becoming in Modern Physics
By Richard T. W. Arthur

NANO-CHIPS 2030
AI On-Chip and Autonomous Chip Systems
Ed. by Boris Murmann, Yiannos Monoli, Bernd Hoefflinger